Genetics and Epigenetics in Systemic Lupus Erythematosus

Chen, Lingyan

Awarding institution:
King's College London

GENETICS AND EPIGENETICS IN SYSTEMIC LUPUS ERYTHEMATOSUS

Lingyan CHEN

Department of Medical & Molecular Genetics

Faculty of Life Sciences & Medicine

King's College London

This thesis is submitted for the degree of

Doctor of Philosophy (PhD)

To my parents

Declaration

The work described in this thesis was carried out within the Vyse Immunogenetics Group in the Department of Medical and Molecular Genetics at King's College London under the supervision of Professor Timothy James Vyse and Dr David Lester Morris and between the years of October 2014 and March 2018. I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Lingyan Chen

March 2018

Abstract

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease associated with a wide range of clinical features involving different organs and the prognosis is also highly variable. It is a complex genetic trait in which more than 80 susceptibility loci show robust genetic association with disease risk.

Firstly, I preformed eQTL mapping, incorporating co-localization analysis of GWAS in the functional immune cells, in order to identify likely causal genes. The results indicate that eQTLs present in a diverse set of immune cells, encompassing both the innate and adaptive immune responses for more than half of the loci. I then integrated genetics, epigenetics and gene expression to delineate the regulatory map of SLE risk loci across human cells and tissues. Using GARFIELD and GoShifter, I demonstrated that SLE genetic associations displayed a marked enrichment pattern for areas of open chromatin in blood, including B cells, T cells, monocytes, and NK cells. Moreover, I found that a large proportion (66%) of the SLE eQTLs showed to overlap with areas of open chromatin, denoting extensive local coordination of genetic influences on gene expression and epigenetics.

SLE patients with renal involvement have more severe clinical outcomes and an increased mortality risk. By calculating the genetic risk score (GRS) using a list of SNPs that are reported to be significantly associated with SLE, I found a significant correlation between GRS and patients with renal involvement – the higher the GRS, the higher probability of getting renal disease. The GRS also correlated inversely with age of SLE onset.

As part of the functional study of post-GWAS, I investigated the protein expression of selected SLE-susceptibility gene products, namely Ikaros family members and OX40L, in a range of immune cells, using multi-parameter flow cytometry. The results reveal some cellular specificity in gene expression in disease. In particular, IKZF3 expression on activated regulatory T cell subsets was decreased, while OX40L expression on B cell subsets was increased in SLE.

In summary, I combined eQTLs and epigenomes to identify the functional tissues and causal genes. In addition, I measured the expression of OX40L and Ikraos family to help understand the cell effects at protein level. Finally, I found that the genetic risk factors that influence the severity of SLE are through a quantitative way but not qualitative way, suggesting that the GRS approach may become a useful factor in predicting outcome in this clinically heterogeneous disease.

Acknowledgements

I would firstly like to express my utmost gratitude to my supervisors Professor Tim Vyse and Dr. David Morris for giving me the opportunity to undertake this PhD and for their very helpful guidance, teaching and encouragement over the last four years. Their support has been invaluable, so thank you a million. It has been a great pleasure to work with them. I am also very grateful to China Scholarship Council (CSC) for funding this PhD Scholarship.

I thank my collaborators in the Professor Julian Knight's Group in Oxford University, Dr. Phe De Jager's Group in Broad Institute, and Dr. Kerrin Small's Group in King's College London. I would like to acknowledge the Biomedical Research Centre Flow Cytometry Core and personally thank Dr. Susanne Heck and her team for their patience and kind suggestions throughout the flow cytometry project. I would like to thank my thesis committee: Dr Kerrin Small, Dr Alan Hodgkinson, and Dr Mike Weale, for their very helpful suggestions throughout my PhD.

This thesis would not have been achievable without the help of my colleagues both past and present in the Immunogenetics group. Thank you so much. In particular, I would like to thank Chris for his encouragement and sincere help everytime I needed. To Pindy, I thank for his very patient teaching in the flow cytometry experiment. To Amy, Ulla, and Andrea, Su, Sarah, and Susan, I thank for all their kind considerations and support throughout my PhD – wishing you all the best for the future. I would like to thank my flatmates – Mengying and Linxi, for providing me a very warm and nice home in London. Thanks to my friends who play badminton with me to help me out of stress.

Thanks to my sister and brothers for their support (especially financially) throughout my PhD. I am so blessed to have such loving and kind family in my life.

Above all else I would like to thank my wonderful parents who have always believed in me and supported my decisions all the time. They are completely special to me. I dedicate this PhD thesis to them.

Publications

Chen, L., Morris, D. L. & Vyse, T. J. ***Genetic advances in systemic lupus erythematosus: an update***. Sep 2017 In : Current Opinion in Rheumatology. 29, 5, p. 423-433

Odhams, C. A., Cortini, A., **Chen, L**., Roberts, A. L., Viñuela, A., Buil, A., Small, K. S., Dermitzakis, E. T., Morris, D. L., Vyse, T. J. & Cunninghame Graham, D. S. ***Mapping eQTLs with RNA-Seq Reveals Novel Susceptibility Genes, Non-Coding RNAs, and Alternative-Splicing Events in Systemic Lupus Erythematosus***. 5 Jan 2017 In : Human Molecular Genetics. 26, 5, p. 1003-1017

Morris, D. L, Sheng, Y., Zhang, Y. , Wang, Y. F. , Zhu, Z. , Tombleson, P. , **Chen, L.** , Cunninghame Graham, D. S. , Bentham, J. , Roberts, A. L. , Chen, R. , Zuo, X. , Wang, T. , Wen, L. , Yang, C. , Liu, L. , Yang, L. , Li, F. , Huang, Y. , Yin, X. & 12 others. ***Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus***. 11 Jul 2016 In : Nature Genetics.

Bentham, J., Morris, D. L., Cunninghame Graham, D. S., Pinder, C. L., Tombleson, P., Behrens, T. W., Martín, J., Fairfax, B. P., Knight, J. C**., Chen, L**., Replogle, J., Syvänen, A. C., Rönnblom, L., Graham, R. R., Wither, J. E., Rioux, J. D., Alarcón-Riquelme, M. E. & Vyse, T. J. ***Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus***. 1 Dec 2015 In : Nature Genetics. 47, 12, p. 1457-1464

Table of Contents

Table of Figures

Table of Tables

Abbreviations

Abbreviation	Meaning
AFR	African
ALL	Acute Lymphoblastic Leukemia
AMR	American
ATAC-Seq	Assay for transposase-accessible chromatin with high throughput sequencing
CAGE	Consortium for the Architecture of Gene Expression
ChIP-Seq	Chromatin immunoprecipitation followed by high-throughput sequencing
DC	Dendritic cell
DHS	DNase 1 hypersensitivity site
DNA	Deoxyribonucleic acid
EAS	East Asian
EBV	Epstein-Barr virus
ENCODE	Encyclopedia of DNA elements
eQTL	Expression quantitative trait loci
eSNP	expression single nucleotide polymorphism
EUR	European
FDR	False discovery rate
GARFIELD	GWAS analysis of regulatory or functional information enrichment with LD correction
GoShifter	Genomic Annotation Shifter
GRS	Genetic Risk Score
GTEx	Genotype-Tissue Expression project
GWAS	Genome-wide association study
IBD	Inflammatory Bowel Disease
IFN	Interferon
LN	Lupus Nephritis
MHC	Major histocompatibility complex
mRNA	Messenger RNA
NGS	Next-generation sequencing
OR	Odds ratio
PBC	Primary biliary cirrhosis
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
pDC	plasmacytoid dendritic cell
RNA	Ribonucleic acid
RNA-Seq	RNA-Sequencing
RTC	Regulatory Trait Concordance
SAS	South Asian
SJS	Stevens-Johnson syndrome
SLE	Systemic lupus erythematosus

Abbreviation	Meaning
SNP	Single nucleotide polymorphism
TCR	T-cell receptor
TF	Transcription factor
TLR	Toll-like receptor
Treg	T-regulatory cell
TSS	Transcription start site
UTR	Untranslated region

Chapter 1. Introduction

1.1 Systemic Lupus Erythematosus

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease characterized by a wide spectrum of signs and symptoms varying among affected individuals and can involve many organs and systems, including the skin, joints, kidneys, lungs, central nervous system, and haematopoietic system (Figure 1-1), thus the diagnosis of SLE remains challenging [1]. Marked differences in SLE population incidence and prevalence rates according to ancestry have been observed, being more prevalent in non-Caucasians populations with a significant sex disparity towards women during the years between menarche and menopause [2]. Although the exact cause of SLE is not fully understood, both genetic and environmental factors have been identified to influence the development of SLE. Genetic interactions with environmental factors, such as UV light exposure and Epstein–Barr virus (EBV) infection, might initiate the disease, resulting in dysregulation in both innate and adaptive immune system.

1.1.1 Epidemiology

The overall incidence rates of SLE range from 0.3 - 23.7 per 100,000 person-years whereas prevalence rates vary approximately from 6.5 – 178.0 per 100,000 [2-4]. The incidence and prevalence rates of SLE are varied worldwide due to the genetic, environmental, sociodemographic and methodological issues. Higher rates are observed in women and in non-Caucasians. The female/male ratio ranges from 3/1 in children to about 9/1 in women of child-bearing age [5, 6]. People of Afro-Caribbean, African and Asian origin have a higher incidence than those of white European origin. A UK study of immigrant women revealed a five-times higher prevalence of SLE in Afro-Caribbean and a three-times higher prevalence in African origin than in Caucasian [7]. Moreover, sex, age and ancestry are associated with disease expression in SLE patients. A recent review by Pons-Estel et al showed that males, childhood-onset, and non-Caucasians SLE patients have a more severe disease [8].

1.1.2 Aetiology

The aetiology of SLE is multifactorial and poorly defined, requiring interplay among genetic predisposition, environmental triggers and hormonal factors. Familial aggregation and twin studies showed that the concordance rate in monozygotic twins (24%) is approximately ten fold higher than in dizygotic twins (2%) [9, 10]. A recent study from Taiwan reported that the heritability was 43.9% and the proportion of phenotypic variance explained by shared and non-shared environmental factors was 25.8 and 30.3%, respectively, suggesting non-heritable factors may play a considerable role in disease pathogenesis [11].

The advent and application of Genome-wide association studies (GWAS) has dramatically advanced knowledge of the genetic cause of SLE. Although in rare cases SLE may be caused by the deficiency of a single gene (e.g., the complement component C4) [12], the disease generally results from the combined effect of variants in a large number of loci. Recent advances of SLE genetic studies are summarized in Section 1.2 (Page 26).

Environmental factors, such as EBV infection, UV light exposure and certain drugs, might trigger the disease. EBV infection may contribute to the activation of the innate immune system and B cell differentiation, leading to the stimulation of autoantibodies production [13]. Exposure to UV light, which might induce DNA breaks and lead to apoptosis, is one of the major factors known to promote the pathogenesis of SLE [14]. Certain drugs, like hydralazine, might trigger SLE through their effects on T cell DNA methylation [15, 16]. In addition, tobacco smoking [17] and silica [18] are found to be potential risk factors for SLE on the basis of their capacity to promote immune response.

The causes of the sexual dimorphism in SLE remain largely unknown. The effect of exogenous hormones on the onset and flares of SLE suggests that hormonal factors represent a component in the pathogenesis of the disease [19]. The risk of SLE development in males is less than females of prepubertal age, but the skewing is less marked than in adults. Of interest, SLE is more common in males with Klinefelter syndrome, characterizing by a 47,XXY genotype, further supporting the limitations of the hormonal hypothesis and raising the possibility of a sex chromosome dose effect [20].

1.1.3 Pathophysiology

SLE is an autoimmune disease characterised by the overproduction of nuclear autoantibodies and proteins. Although the exact mechanism of autoimmune responses is not fully understood,

both innate and adaptive immune disturbances occur in SLE. SLE is widely thought to be a self-immunization disease driven by defective clearance of dead cells. The accumulation of apoptotic cells in various tissues and the signals released by them act as potential direct triggers of innate immune system activation. Nucleic acid-containing immune complex and cytoplasmic RNA and DNA are potential stimuli for the activation of nucleic acid-responsive endosomal TLRs and TLR-independent nucleic acid sensors, leading to the production of type I interferons (IFNs) [21].

Type I IFNs not only activate dendritic cells (DCs) and nature killer (NK) cells involved in innate immunity, but also activate B cells and T cells to trigger the adaptive immune system. T cells are recognised as the most efficient drivers of B cell differentiation [22]. Although lymphocytopaenia is a typical feature of SLE, specific T cell populations are expanded in SLE patients. The expansion of T follicular helper (Tfh) cells promotes the differentiation of autoantibody-producing B cells [23]. The expansion of CD8+ memory the T cell population is associated with poor prognosis in SLE [24]. Regulatory T (Treg) cells, that have the function of suppressing immune response, are also considered to be an aspect of SLE pathogenesis [25]. However, reports on the numbers and functions of Treg in SLE are controversial. Some studies report that Treg cells are depleted in patients with lupus nephritis [26-30], while other studies observe no alteration or increased level of Treg cells in SLE compared with those in healthy controls [31-33]. B cell function is also altered in SLE, leading to the production of autoantibodies, cytokines and augmented presentation of antigen to T cells. Long-lived plasma cells are proposed sources of anti-Sm and anti-Ro autoantibodies that are refractory to anti-B cell therapy while circulating plasmablasts are sources of anti-dsDNA autoantibodies which are more amenable to B cell depletion therapy [34, 35]. The proposed pathogenesis of SLE is shown in Figure 1-2 [36].

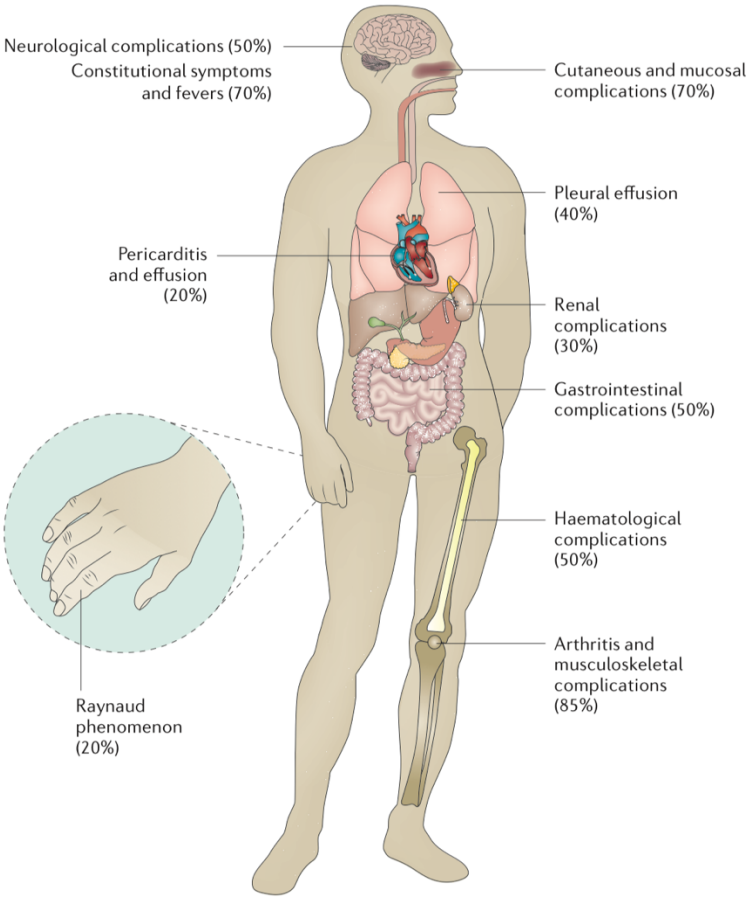


Figure 1- 1. Clinical heterogeneity of SLE.

The multifaceted nature of SLE is shown by the number of different organ systems that can be affected. The average frequency of the most common complications is indicated in parentheses. Adapted from [1].

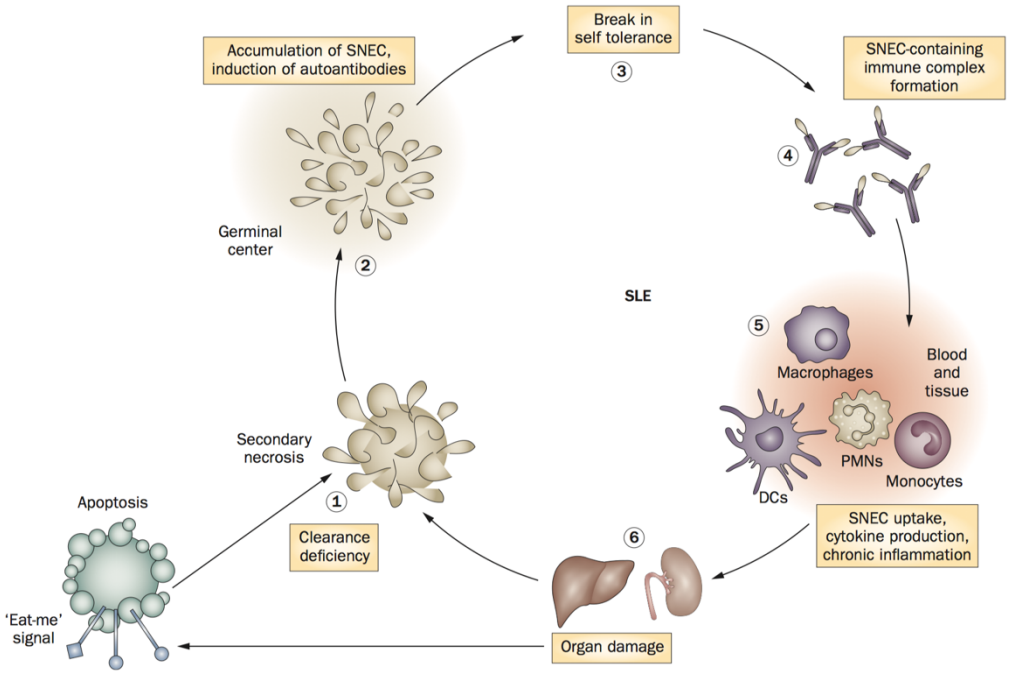


Figure 1- 2. Aetiology and pathogenesis of SLE.

In the scenario of deficient clearance of apoptotic cells, dying cells are prone to progress to secondary necrosis (1), which then exposes autoantigens in an inflammatory context. Undigested apoptotic remnants (SNEC) accumulate in germinal centres (2) and are presented by follicular DCs to autoreactive B cells, breaking tolerance to self. Autoreactive B cells might undergo proliferation and activation, eliciting immune responses against nuclear components (3). Antinuclear autoantibodies encounter nucleic-acid-containing apoptotic remnants either in circulation or deposited in tissues to form immune complexes (4). SNEC-containing immune complexes are cleared by blood-borne phagocytes, macrophages and DCs, which, in turn, secrete high amounts of inflammatory cytokines (5). The final outcome is multiple organ damage and, subsequently, enhanced cell death, closing a vicious circle that leads to the establishment of chronic inflammation (6). Steps 1–3 represent the events involved in the aetiology of SLE; steps 4–6 correspond to the description of the 'lupus pathogen' causing disease manifestations (pathogenesis). Abbreviations: DC, dendritic cell; PMN, polymorphonuclear leukocyte; SLE, systemic lupus erythematosus; SNEC, secondary necrotic cell-derived material. Adapted from reference [36].

1.2 Genetics of SLE

Although the exact aetiology of lupus is not fully understood, a strong genetic link has been identified through the application of family and large-scale genome-wide association studies (GWAS). The concordance rate in monozygotic twins (24%) is approximately ten fold higher than in dizygotic twins (2%) [9, 10]. Lawrence JS et al's study in a European population from a family survey estimated the overall heritability of SLE to be 66 +/- 11% [37]. However, a recent study from Taiwan reported that the heritability was 43.9% (with 25.8% for shared environmental factors), suggesting non-heritable factors may play a considerable role in disease pathogenesis [11].

1.2.1 Human genome variations

The human genome contains various forms of variation and the most common one is single nucleotide polymorphisms (SNP) -- variants in an organism's DNA sequence where more than one allele is present at a detectable frequency in the population. SNPs can be classified into different groups depending on their positions and effects. For example, a synonymous SNP is one located in a protein-coding region but doesn't affect the amino acid sequence after translation. Other types of genetic variants include small indels (insertion and/or deletion) and structural variants. A brief summary of the different types of human variations are summarised in Figure 1-3. In this thesis, only the effects of SNPs are evaluated.

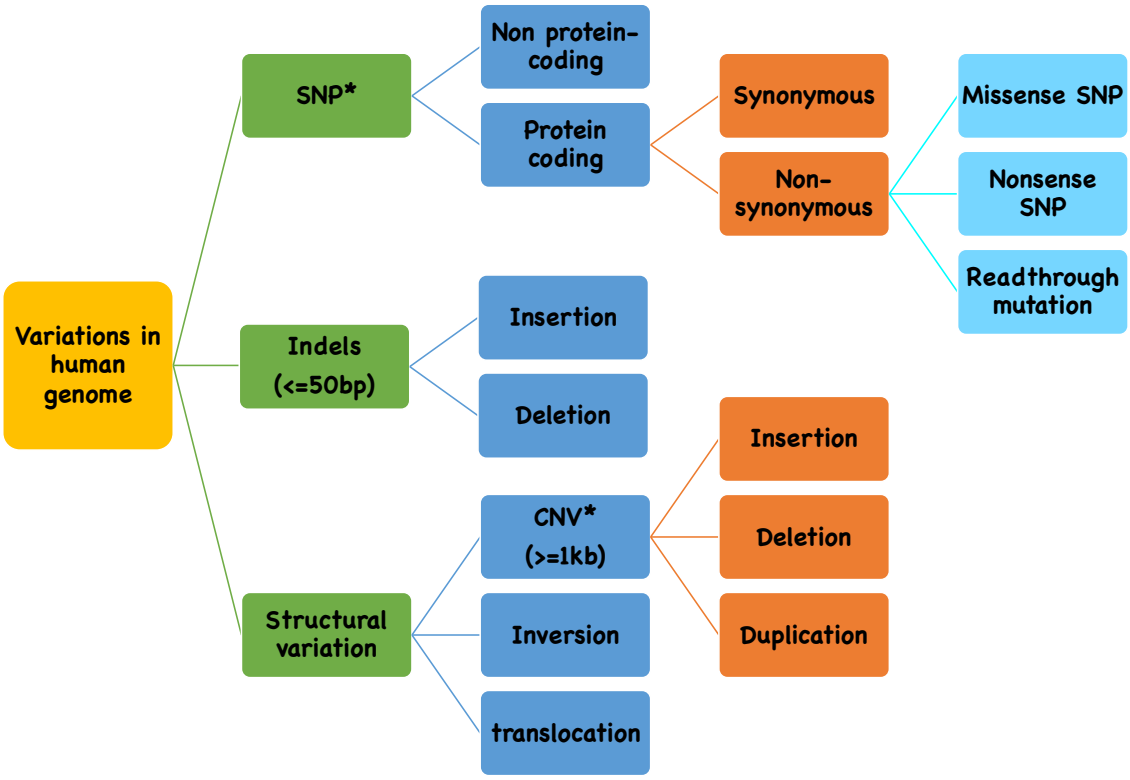


Figure 1- 3. Variations in human genome.
*SNP: Single nucleotide polymorphism; **CNV: copy number variation

1.2.2 Insights from SLE GWAS

As mentioned above, SLE is a disease with a highly genetic background but does not follow the Mendelian patterns of inheritance, and so it is termed a complex or non-Mendelian trait. Complex traits are multi-factorial with both genetic and environmental contributions. In addition, there is no single locus containing alleles that are sufficient for occurrence of disease in most cases. Genome-wide association studies (GWAS) are used to investigate the genetic basis of a disease by comparing a group of patients with disease to a group of healthy controls. The advent and application of GWAS has dramatically advanced knowledge of the genetic aetiology of SLE.

There have been seven SLE GWAS in the European population [38-44], six Asian GWAS [45-50], and one GWAS of Amerindian ancestry [51], as well as subsequent meta-analyses and large-scale replication studies [52-55], published since 2008.

Currently, 84 genetic loci are implicated as SLE risk (Figure 1-4: The CIRCOS plot and Table 1-1), which, in order to avoid likely spurious associations, includes genetic associations with a p-value less than 5×10^{-8} tested in a total sample size of at least 1000 individuals. The interactive version of a continually updated resource with details on SLE associations can be access through

the following link: http://insidegen.com/insidegen-LUPUS-Associations.html. Major findings from GWASs published in the last two years are summarized below.

European GWAS

The largest European GWAS of SLE conducted by our group [43], comprising 7,219 SLE cases and 15,991 controls of European decent, provided considerable power to detect disease risk loci. Notably, the study identified 43 susceptibility loci, ten of which were novel loci: *SPRED2, IKZF2, IL12A, TCF7-SKP1, DHCR7-NADSYN1, SH2B3, RAD51B, CIITA-SOCS1, PLD2*, and *CXorf21*. One of the great challenges posed by interpreting GWAS data is determining the causal genes implicated by the genetic association data. As will be discussed in Section 1.3, my work makes a considerable contribution and not trivial effort in naming the genes in the above list (Table 1-1). Irrespective of the underlying causal genes, the European GWAS paper concluded that the heritability explained by the risk alleles mapped at these loci was 15.3%, which is a large increase over the 8.7% reported by So et al in 2011 using the same measure [56].

Asian GWAS

An extensive large-scale fine mapping study using the Immunochip conducted in 4,478 SLE cases and 12,656 controls from six East Asian cohorts identified 10 novel loci [50] in Asians, encompassing *GTF2IRD1-GTF2I, DEF6, IL12B, TCF7, TERT, CD226, PCNXL3, RASGRP1, SYNGR1,* and *SIGLEC6*. Some of these were previously reported to be associated in Europeans, for example, *DEF6* and *TCF7*. The identification of these risk loci increased the explained heritability to 24% in Asian SLE. Nevertheless, the Immunochip was designed from predominantly European genetic information and will therefore be less informative and not as representative of the genetic variation in Asian populations as it is in Europeans [57].

Trans ancestry meta analyses of GWAS

A comparison of genetic association signals across the genome in European and Asian populations suggested that SLE susceptibility loci were shared extensively between both populations [54]. This motivated a trans ancestral approach at the genome-wide level to provide evidence of shared genetic components in the two populations and a search for additional SLE associated loci. The study by Morris and Sheng et al [54], that combined three GWAS from two ethnicities: two Chinese (a total of 1659 cases and 3,398 controls) and one European (4,044

cases + 6,959 controls), found evidence of considerable commonality in terms of SLE association signals as well as mapping novel susceptibility loci, including *CD45, IKBKE, LBH, LPP-TPRG1-AS1, ATXN1, BACH2, GTF2I, JAK2, RNASEH2C*, and *ZFP90*. Notably, this study suggested that the increased prevalence of SLE in non-European (including Asians) has a genetic basis by a comparison of genetic risk scores (GRS) between populations. Moreover, by using all genotyped SNPs (DNA chip) to calculate heritability explained, the explained variation (Vg) increased to 28% in Chinese subjects and 27% Europeans using the GCTA algorithm [58] . While there are still some uncertainties in the methodology for calculating heritability explained, this shows very strong evidence that we are making progress on the understanding of SLE heritability. The latest large-scale trans ancestral study using Immunochip [59], comprising three ancestries: European (EA: 6,748,cases and 11,516 controls), African-American (AA: 2,970 case and 2,452 controls), and Hispanic Amerindian (HA: 1,872 cases and 2,016 controls), identified eight novel loci for EA (*TMEM39A-TIMMDC1, DGKQ, LRRC16A, SLC17A4, OLIG3-LOC100130476, FAM86B3P, PKIA-ZC2HC1A,* and *GRB2*), two for AA (*PTTG1-MIR146A* and *PLAT*) and two for HA (*GALC* and *CLEC16A*). By comparing results across different populations, both ancestry-dependent and ancestry-independent contributions to SLE risk were identified with the caveat of unequal cohort sizes. The study reveals further evidence of sharing of genetic risk loci between ancestries as well as evidence that each individual population carries unique genetic risk factors at the locus level and at the allelic level.

Table 1- 1. A summary of SLE risk loci

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
1	1p36.11	rs4649203	24519920	G	intergenic	*IL28RA*	Chinese	24070858
2	1p13.2	rs2476601	114377568	A	missense	*PTPN22*	American and European	19838195, 26502338
3	1q23.3	rs1801274	161479745	C	missense	*FCGR2A*	European	26502338
4	1q25.1	rs1234315	173178463	A	intergenic	*TNFSF4*	Asian and European	19838193, 22820624
		rs2205960	173191475	T	intergenic		Asian, European-American, African-American and Hispanic	19838193, 19838195, 22820624, 23874208
		rs704840	173226195	G	intergenic		European	26502338
		rs1418190	173361979	G	intergenic		Chinese	25890262
		rs4916219	173373183	A	intergenic		Chinese	25890262
5	1q25.3	rs17849502	183532580	T	missense	*NCF2*	European-American, African-American, Hispanic and Korean	24163247
		rs35937854	183536089	G	missense		European-American, African-American, Hispanic and Korean	24163247
		rs10911363	183549757	T	intron		UK, American and Swedish	22046141
		rs10911628	184649503	NA	intergenic	*EDEM3*	European	24871463
6	1q31.3	rs34889541	198594769	G	intergenic	*CD45*	Chinese and European	27399966
7	1q32.1	rs2297550	206643772	G	5'-UTR	*IKBKE*	Chinese and European	27399966
		rs3024505	206939904	A	intergenic	*IL10*	American and European	19838195, 26502338
8	1q42.3	rs9782955	236039877	C	intron	*LYST*	European	26502338
9	2p23.1	rs7579944	30445026	C	intergenic	*LBH*	Chinese and European	27399966

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
		rs17321999	30479857	C	intron		Chinese and European	27399966
10	2p22.3	rs13385731	33701890	G	intron	*RASGRP3*	Chinese	19838193
11	2p14	rs6740462	65667272	A	intron	*SPRED2*	European	26502338
12	2p13.1	rs4852324	74202578	T	intron	*DGUOK-TET3*	Asian	23273568
		rs6705628	74208362	C	non coding		Asian	23273568
13	2q24.2	rs2111485	163110536	G	intergenic	*IFIH1*	European	26502338
		rs10930046	163137983	G	missense		African and European	23441136
14	2q32.3	rs3821236	191902758	NA	intron	*STAT4*	Korean	23740238
		rs7601754	191940451	NA	intron		Asian	19225526
		rs11889341	191943742	T	intron		European	26502338
		rs10931481	191954852	T	intron		European	23053960
		rs6736175	191946322	C	intron		European	26502338
15	2q34	rs3768792	213871709	C	3'-UTR	*IKZF2*	European	26502338
16	3p14.3	rs6445972	58321707	C	intron	*ABHD6-PXK*	European Americans, African Americans, Asian Americans and Hispanic Americans	25620976
		rs6445975	58370177	C	intron		European	18204446
		rs9311676	58470351	C	intergenic		European	26502338
17	3q13.33	rs1132200	119150836	G	missense	*TMEM39A*	European, African American and Asian	22464253, 25890262
		rs1131265	119222456	G	synonymous		European	28714469
		rs6804441	119260944	A	intron	*CD80*	Asian	23273568
		rs2222631	119272391	G	intron		Asian	25862617
18	3q25.33	rs564799	159728987	C	intron	*IL12A*	European	26502338
19	3q26.2	rs10936599	169492101	C	synonymous	*MYNN*	Asian and European	28108556
20	3q28	rs6762714	188470238	T	intron	*LPP-TPRG1-AS1*	Chinese and European	27399966
21	4p16	rs4690229	970724	T	3'-UTR	*DGKQ*	European	28714469
22	4q21.3	rs340630	87958395	A	intron	*AFF1*	Japanese	22291604
23	4q24	rs10028805	102737250	G	intron	*BANK1*	European	26502338
		rs17266594	102750922	C	intron		Chinese	19357697
24	4q27	rs907715	123535053	G	intron	*IL21*	European and African American	21425124

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
25	5p15.33	rs7726159	1282319	A	intron	*TERT*	Asian	26808113
26	5q31.1	rs7726414	133431834	T	intergenic	*TCF7-SKP1*	European and Asian	26502338, 26808113
27	5q33.1	rs7708392	150457485	C	intron	*TNIP1*	American and Swedish	19838195
28	5q33.3	rs2421184	159459931	A	intron	*IL12B*	Asian	26808113
		rs2431697	159879978	A	intergenic	*PTTG1-MIR146A*	Chinese and European	25890262, 26502338
		rs2431697	159879978	A	intergenic		African American	28714469
29	6p23	rs17603856	16630898	T	intron	*ATXN1*	Chinese and European	27399966
		rs2327832	137973068	C	intron	*OLIG3-LOC100130476*	European	28714469
30	6p22	rs36014129	25884519	A	intron	*SLC17A4*	European	28714469
		rs35789010	25514179	A	intron	*LRRC16A*	European	28714469
31	6p21.32-33	rs3906272	31262924	A	intron	*HLA-C-HLA-B*	UK, Spanish and Filipino	22233601
		rs2246618	31478986	A	upstream	*MICB*	European	23084292
		rs1269852	32080191	A	intron	*MSH5-SAPCD1-TNXB-ATF6B*	European	18204446, 19838195
		rs8192591	32185796	G	missense	*NOTCH4*	European	23084292
		rs9270984	32573991	T	intron	*HLA-DRB1*	Asian and European	23273568, 20169177, 22233601
		rs2187668	32605884	A	intron	*HLA-DQA1*	European	18204098
		rs9275572	32678999	NA	intergenic	*HLAregion*	European	18204446
		rs2051549	32730086	A	intron	*HLA-DQA2*	European	23053960
32	6p21.31	rs3748079	33588147	C	5'-UTR	*ITPR3*	Japanese	18219441
		rs3734266	34823187	G	intron	*UHRF1BP1*	Asian and European	23273568, 26502338
		rs11755393	34824636	G	missense		American and Swedish	19838195
		rs2762340	35052635	G	intron	*ANKS1A*	Asian and European	28108556
		rs10807150	35304497	C	intron	*DEF6*	Asian	26808113
33	6q15	rs597325	91002494	G	intron	*BACH2*	Chinese and European	27399966
34	6q21	rs548234	106568034	G	intergenic		Chinese	19838193

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
		rs6568431	106588806	A	intergenic	*PRDM1-ATG5*	American and European	19838195, 26502338
35	6q23	rs6932056	138242437	C	intergenic	*TNFAIP3*	European	26502338
		rs2327832	137973068	C	intron	*OLIG3-LOC100130476*	European	28714469
36	7p15.1	rs849142	28185891	T	intron	*JAZF1*	American and European	19838195, 26502338
37	7p12.2	rs2366293	50227828	G	intergenic	*IKZF1*	UK, American and Swedish	22046141
		rs4917014	50305863	A	intergenic		Chinese and European	19838193, 26502338
38	7q11.23	rs73135369	73940978	C	intron	*GTF2IRD1-GTF2I*	Chinese and European	27399966
		rs73366469	74619286	C	intergenic		Asian and European	26808113
		rs1167796	75173180	G	intron	*HIP1*	Chinese	19838193
		rs6964720	75180344	NA	intron		Asian	26663301
39	7q32.1	rs729302	128568960	A	intergenic	*TNPO3-IRF5*	Asian and European	23273568, 23053960
		rs4728142	128573967	A	intergenic		Chinese and European	19838193, 24871463
		rs2070197	128589000	C	3'-UTR		American and Swedish	19838195
		rs12537284	128717906	A	intergenic		European	18204446
40	8p11	rs1804182	42033519	A	nonsense	*PLAT*	African American	28714469
41	8p23.1	rs2980512	8140901	C	intergenic	*FAM86B3P*	European	28714469
		rs6985109	10767748	A	intron	*XKR6*	European	18204446
		rs7836059	11272164	T	intron	*C8orf12*	European	18204446
		rs12680762	11332026	A	intron	*FAM167A*	European	23053960
		rs7812879	11340181	A	intergenic	*BLK*	Chinese	19838193
		rs2736340	11343973	T	intergenic		American and Swedish	19838195
		rs2248932	11391650	A	intron		Chinese	19838193
42	8q12	rs7829816	56849386	C	intron	*LYN*	European	18204446
43	8q21	rs1966115	79556891	A	intergenic	*PKIA-ZC2HC1A*	European	28714469

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
44	9p24	rs1887428	4984530	G	upstream	*JAK2*	Chinese and European	27399966
45	10q11.23	rs7097397	50025396	G	missense	*WDFY4*	Asian	20169177
		rs877819	50042951	A	intron		Asian and European	23273568, 26502338
		rs1913517	50119054	A	intron		Chinese	19838193
46	10q21.2	rs4948496	63805617	C	intron	*ARID5B*	Asian and European	23273568, 26502338
47	10q24.33	rs4917385	103243964	NA	intergenic	*USMG5-miR1307*	American	26606652
48	11p15.5	rs12802200	566936	C	intron	*IRF7-PHRF1*	European	26502338
		rs4963128	589564	C	intron		American and European	19838195, 18204446
49	11p13	rs2732552	35084592	T	intergenic	*CD44*	European, African American and Asian	21194677, 26502338
50	11q13.1	rs1308020	65497558	T	intergenic	*RNASEH2C*	Asian and European	28108556
		rs494003	65542298	A	3'-UTR		Chinese and European	27399966
		rs2009453	65632057	C	intron	*PCNXL3*	Asian	26808113
51	11q13.4	rs3794060	71187679	C	intron	*DHCR7-NADSYN1*	European	26502338
		rs11235604	72533536	T	missense	*ATG16L2*	Asian and European	28108556
		rs11235667	72863697	NA	intergenic	*FCHSD2-P2RY2*	Asian	26663301
52	11q23.3	rs11603023	118486067	T	intron	*PHLDB1*	Asian	24001599
		rs4639966	118573519	G	downstream		Chinese	19838193
		rs10892301	118735476	G	intergenic	*CXCR5*	Asian	24001599
53	11q24.3	rs6590330	128311059	A	intergenic	*ETS1-FLI1*	Asian	19838193, 23273568
		rs7941765	128499000	C	downstream		European	26502338
54	12p13	rs12822507	12773521	A	intron	*CREBL2*	Asian	23273568
		rs10845606	12834894	C	intron	*GPR19*	Asian	23273568
		rs34330	12870695	C	5'-UTR	*CDKN1B*	Asian	23273568
55	12q12	rs10506216	43130885	T	intergenic	*PRICKLE1*	European	26316170
56	12q23.2	rs4622329	102321935	A	intron	*DRAM1*	Asian	23273568
57	12q24.12	rs10774625	111910219	A	intron	*SH2B3-ATXN2*	European	26502338

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
58	12q24.32	rs1059312	129278864	C	synonymous	*SLC15A4*	European	26502338
		rs10847697	129299385	A	synonymous		Chinese	19838193
59	13q14.11	rs7329174	41558110	G	intron	*ELF1*	Asian	21044949
60	14q13.2	rs8016947	35832666	A	intron	*NFKBIA*	Chinese	24070858
61	14q24.1	rs4902562	68731458	A	intron	*RAD51B*	European	26502338
62	14q31	rs11845506	88383035	C	intron	*GALC*	Hispanic	28714469
63	15q14	rs12900339	38635185	A	intergenic	*RASGRP1*	Asian	26808113
		rs11073328	38764843	NA	intron	*FAM98B*	European	24871463
64	15q24	rs34933034	75079474	NA	intron	*CSK*	European	23042117
		rs2289583	75311036	A	intron		European	26502338
65	15q26.2	rs8023715	97607681	NA	intergenic	*SPATA8*	European	24871463
66	16p13	rs8054198	11038360	C	5'-UTR	*CLEC16A -CIITA-SOCS1*	Hispanic	28714469
		rs9652601	11174365	G	intron		European	26502338
		rs12599402	11189888	A	intron		Chinese	20805369
67	16p11.2	rs16972959	23901376	A	intron	*PRKCB*	Chinese	21134959
		rs7197475	30642867	A	intergenic	*ITGAM-ITGAX*	Chinese	19838193
		rs34572943	31272353	A	intron		European	26502338, 18204446
		rs11574637	31368874	C	intron		European	18204098
68	16q13	rs223881	57386566	C	intergenic	*CCL22*	Asian and European	28108556
69	16q22.1	rs1170426	68603798	C	intron	*ZFP90*	Chinese and European	27399966
70	16q24.1	rs2934498	85968282	G	intergenic	*IRF8*	Chinese	25890262
		rs11644034	85972612	G	intergenic		European, African American and Asian	22464253, 26502338
		rs2280381	86018633	A	intergenic		UK, American and Swedish	22046141
71	17p13.2	rs2286672	4712617	T	missense	*PLD2*	European	26502338
72	17q12	rs2941509	37921194	A	3'-UTR	*IKZF3*	European	26502338
73	17q25	rs930297	73404537	A	intergenic	*GRB2*	European	28714469
74	18q22.2	rs1610555	69875911	T	intron	*CD226*	Asian	26808113
75	19p13	rs3093030	10397403	A	downstream	*ICAM1-ICAM4-ICAM5*	European, African, Hispanic, and Korean	22523428
		rs280519	10472933	A	intron	*TYK2*	UK, American and Swedish	22046141
		rs2304256	10475652	C	missense		European	26502338

No.	Risk Region	Lead SNP in Publication	Position (GRCh37)	Risk allele	dbSNP Function	Locus Gene*	Population	References (PUBMEDID)
76	19q13.41	rs2305772	51530488	G	missense	*SIGLEC6*	Asian	26808113
77	20q13.12	rs4810485	44747947	T	intron	*CD40*	Greek and Turkey	21914625
78	20q13.13	rs11697848	48575315	NA	intergenic	*RNF114*	European	24871463
79	22q11.21	rs463426	21809185	A	intergenic	*UBE2L3-YDJC-HIC2*	Chinese	19838193
		rs131654	21917190	A	intron		Chinese	19838193
		rs7444	21976934	C	3'-UTR		European	26502338
80	22q13.1	rs61616683	39359768	T	intron	*SYNGR1*	Asian	26808113
81	Xp22.2	rs7062536	12839152	A	intron	*PRPS2*	Asian	25149475
		rs3853839	12907658	G	3'-UTR	*TLR7*	Asian	20733074
82	Xp21.2	rs887369	30577846	C	synonymous	*CXorf21*	European	26502338
83	Xp11.21	rs5914778	56758231	NA	intron	*LINC01420*	Chinese	26635088
84	Xq28	rs17422	153227426	C	intron	*NAA10-HCFC1-TMEM187*	European Americans, African Americans, Asian Americans and Hispanic Americans	25149475, 19329491
		rs1734787	153325446	C	intron	*IRAK1-MECP2*	European	26502338

** Locus gene, is the gene that located close to the GWAS SNP, or the gene that annotated in the corresponding publications.

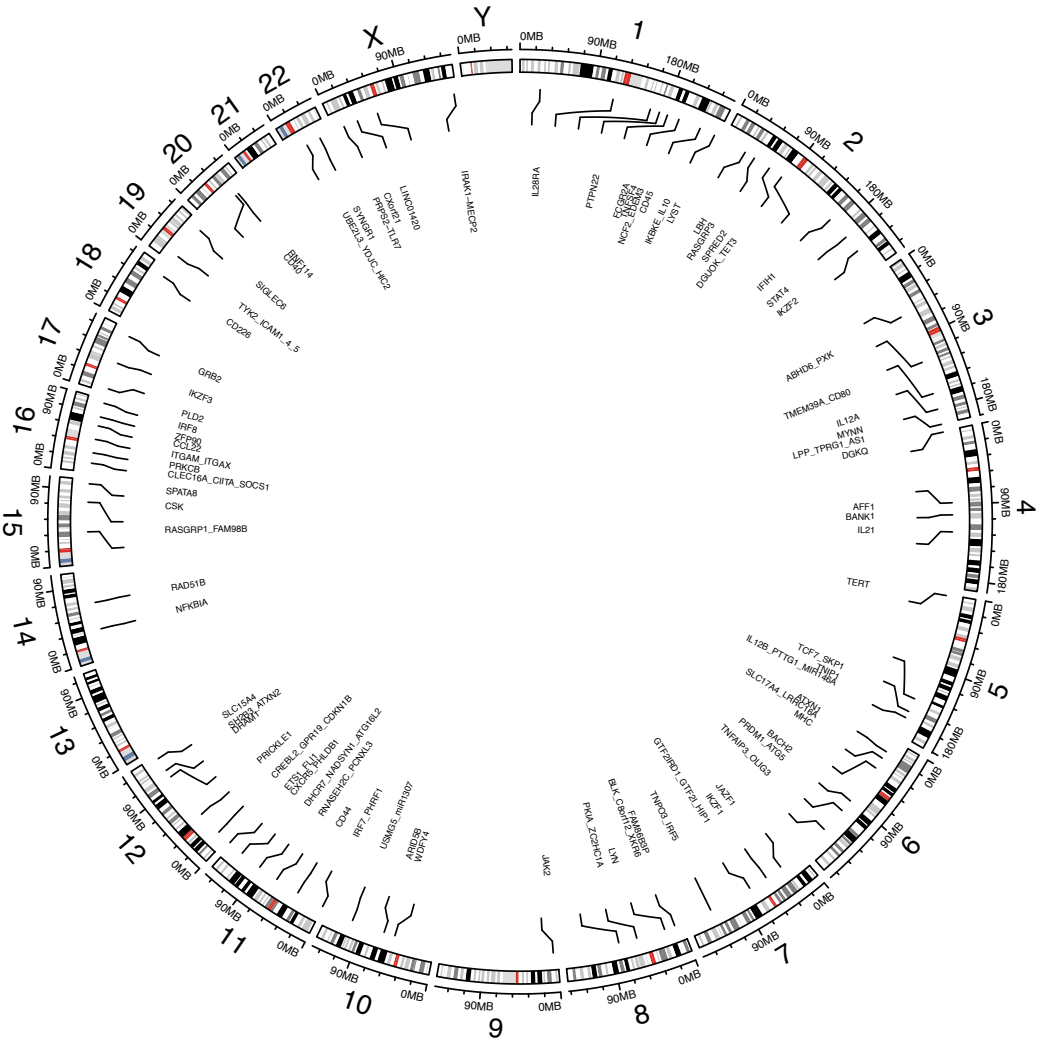


Figure 1- 4. SLE risk loci in genomic context

The CIRCOS plot shows genes located within the SLE risk loci (84 in total) according to their genomic position. The full list of variants and locus genes for this plot is summarized in Table 1-1. The red block in each chromosome indicates the centromere of the chromosome. Each chromosome arm is divided into cytogenetic bands of hg19.

1.2.3 Missing heritability

In summary, the chip heritability identified by the latest GWAS have explained around 28% of the disease heritability: a marked improvement on 8.3% calculated in 2011 [56]. Although the overall heritability of complex disease is complicated to estimate, a study in a European population from a family survey did estimate a heritability of 66 +/- 11% for SLE [37], indicating there is more than 50% missing heritability in SLE from current GWASs. If we assume that the total heritability is

43.9% (with 25.8% for shared environmental factors) estimated from a Taiwanese population [11], there is still one third of heritability left to explain. Explanations for the missing heritability, including larger numbers of variants of smaller effect, rarer variants (possibly with larger effects) that are not present on genotyping arrays or structural variants poorly captured by existing arrays, as well as epigenetic modifications, have been suggested [60]. Innovations in genotyping and sequencing technologies, like the Immunochip platform [50, 59], and next generation sequencing (NGS, as described below) will advance the investigation into common and rare variants and potential effects on the immune system, enhancing our understanding of the genetic risk of SLE. The linkage disequilibrium (LD) that exists in the human genome facilitates the mapping of risk loci by reducing the number of genetic variants required for GWAS; however, the same correlation between genetic polymorphisms at these susceptibility loci then obscures attempts to identify the actual causal allele(s) at risk loci. Bayesian fine mapping approaches had been proposed to derive smaller sets of SNPs (termed 'credibility sets') as the most likely causal variants at risk loci [61]. Nevertheless, statistical methods are inadequate to fully resolve the problem caused by LD. In order to further pursue likely causal SNPs within any given credibility set, mapping across ancestries can be applied to narrow down the target set of SNPs and the functional effect of SNPs can be studied in silico. As the majority of variants within causal credibility sets are non-coding [62, 63], function is inferred using gene transcript expression data and epigenetic modification data (as described in Section 1.3).

1.3 Functional annotation of genetic variants

More than 80 susceptibility loci are now reported to show robust genetic association with SLE. Nevertheless, the overwhelming majority of these loci (~90%) lie in noncoding regions of the genome [64]. In the post-GWAS era, a major challenge is linking GWAS findings to additional molecular function in order to better understand the identity of the genes that underlie the genetic associations. It is essential to bridge the gap between genotype and function in order to fully exploit the potential of GWAS. Otherwise, the results from GWAS are essentially a list of genomic positions. Furthermore, identification of molecular intermediaries, such as RNAs or proteins, between genotype and disease provides practical implications for developing pharmacological interventions.

1.3.1 Incorporating gene expression profiling to infer causal genes

It is now possible to investigate the effect of risk variants on level of intermediate molecules, for example, gene expression, methylation, protein abundance or metabolite levels, due to the advances in a variety of ~omics techniques. Genetical genomics, which combines genome-wide expression profiling and genome-wide marker-based genotyping to members from a segregating population, has been proposed for discovering regulatory pathways [65]. This strategy takes advantage of the heritability of gene expression profiles to identify genetic variants that correlate with changes in gene expression [66-68].

Expression quantitative trait loci (eQTLs)

An eQTL (expression quantitative trait locus) is a genetic element associated with the expression level of a given gene. For the purpose of this thesis, the genetic element is a SNP and the expression level (mRNA abundance) of a given gene is considered as an intermediate molecular trait, i.e. 'expression quantitative trait'. As exemplified by the eQTL mapping literature, regulatory polymorphisms can be classified as "in cis" or "in trans" on the basis of their physical distance from the regulated gene. Regulatory variation that is in the vicinity of the target gene is classified as being cis, while trans-eQTLs are SNP-gene pairs that are further away from each other (Figure 1-5). A cis association is often defined as the variant being present in a ±1Mb window surrounding the transcription start site (TSS) of the gene, and trans effects are defined as associations involving SNPs elsewhere in the genome [69].

Comprehensive eQTL mapping can provide evidence for interpreting GWAS associations and constructing potential genes interaction networks [65]. Recent research has found that over 40% of trait-associated SNPs are more likely to have a cis-acting effect on gene expression [70]. Another important potential advantage of studying eQTLs is that they can provide insight into the disease mechanism and underlying pathways. The simple integration of trait association with gene expression data in disease relevant cell types has indicated the probable causal gene(s) in a number of instances. For example, variants that associated with asthma, locating across a large block (200kb) encompassing 19 genes, were found to increase the expression of a single gene in LCL: *ORMDL3* (Sphingolipid Biosynthesis Regulator 3) [71]. The expression of *ORMDL3* was found to associate with the abnormalities in endoplasmic reticulum-mediated calcium signalling and modulating the inflammatory responses [72].

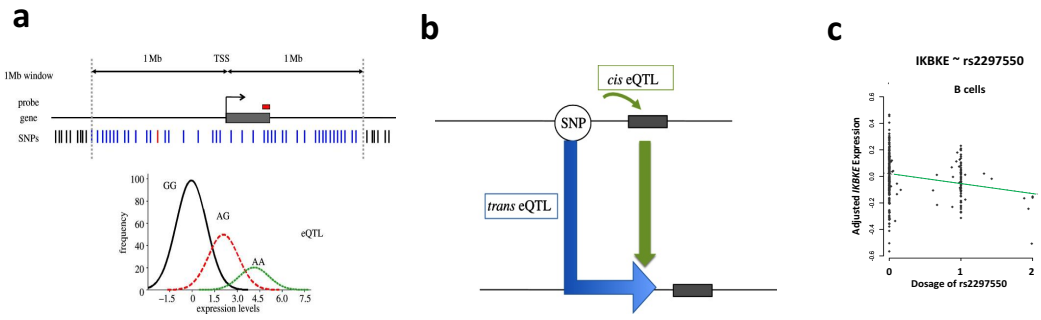


Figure 1- 5**. eQTL Mapping.**

(a) A typical eQTL - many SNPs tested against levels of expression measured by a probe or by other means. The panel below illustrates the difference in distributions of expression values stratified by the SNP genotypes of the most significant SNP. (b) eQTLs can be classified according to their location (*cis*- or *trans*-) [69]. (c) eQTL mapping with linear regression. Genotype versus *IKBKE* (Inducible IkappaB Kinase) adjusted expression level in B cells, with the estimated regression line shown in green. The genotypes are codes as 0~2, termed 'dosage', counting the number of reference alleles. In this case, the dosage is counted on allele G, so genotypes of AA, AG and GG are coded as 0,1 and 2, respectively. Data points that are not integer are obtained from imputation, arising from the probabilities of each genotype. Data for (c) are obtained from Fairfax et al [73].

Application of eQTL mapping to GWAS results

Assisted by dense genome coverage of the reference panel from the 1000 Genome project [74], imputation and Bayesian inference provided evidence for missense variants underpinning association for eight genes, including *PTPN22, FCGR2A, NCF2, IFIH1, WDFY4, ITGAM, PLD2*, and *TYK2* [43]. However, as mentioned above, the majority (~ 90%) of disease associated loci in SLE are located outside of protein-coding regions, and so might exert their function through altering gene expression rather than by altering protein structure. Of note, an over-representation (n=16) of transcription factors among the 43 SLE susceptibility genes have been annotated in our recent European GWAS [43], further indicating that perturbed gene regulation is a major functional risk factor for SLE. eQTL mapping, which combines genome-wide expression profiling and genome-wide marker-based genotyping, takes advantage of the heritability of gene expression profiles to identify genetic variants that correlate with changes in gene expression.

Some studies [50, 55] use public databases, such as the whole blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) [75], and tissue-specific GTEx portal (http://www.gtexportal.org/home/) [76], to determine whether a disease-associated SNP is a significant eQTL. Of note, limitations exist when applying eQTL analysis to the GTEx whole blood data sets; for example, we seek eQTL in specific immune cell subsets when studying autoimmune disease.

In order to highlight potential causal genes at the susceptibility loci robustly, it is essential to integrate the disease association and eQTL data using a co-localisation approach. That is, to establish that the same genetic variants that underlie the disease association also underlie the eQTL. The presence of LD in the genome can readily confuse this overlap. Co-localisation methods, like the regulatory trait concordance (RTC) [77], conditional analysis [58], and Bayesian co-localisation [78] can be employed to infer that the disease association and eQTL have the same allelic basis. As many variants have weak eQTL effects, erroneous conclusions will be made if analyses for co-localisation are not performed. Many SNPs may be observed to be both eQTLs, even if weak, and also associated with the target phenotype (e.g. disease), so it is imperative to separate out a spurious eQTL overlap from ones that co-localise and hence makes causality more likely. Figure 1-6 shows an example for the illustration of co-localisation results.

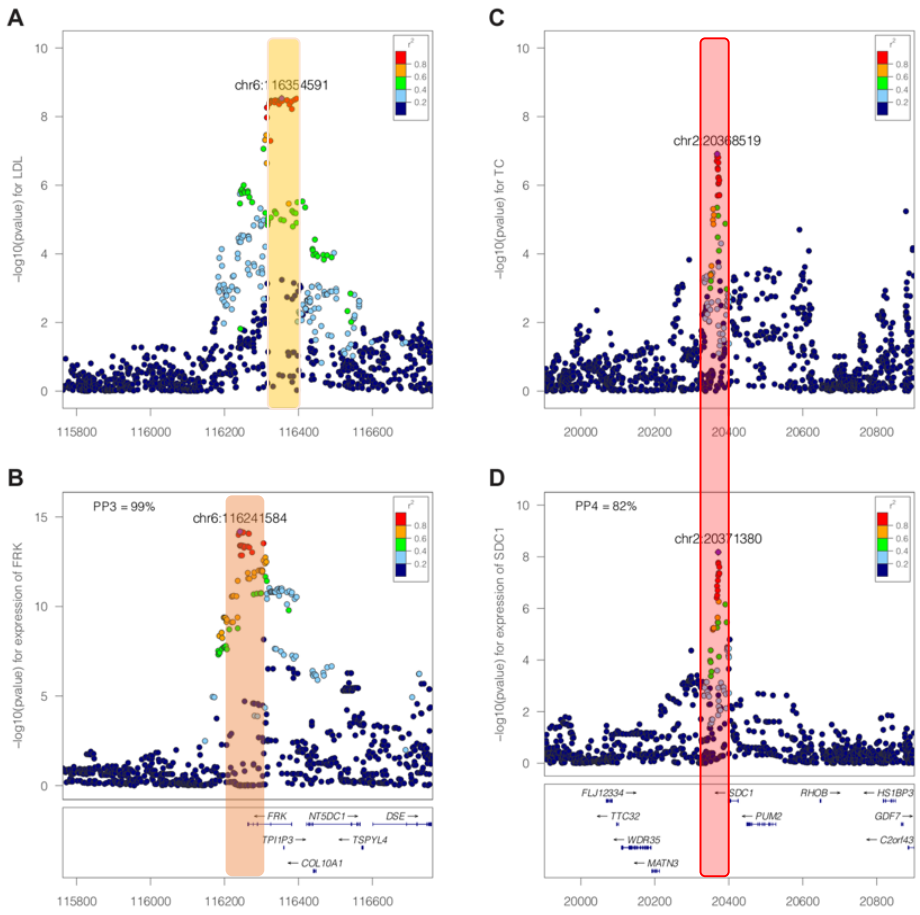


Figure 1- 6**. Illustration of co-localisation results.**
Left panel (A–B, *FRK* gene and LDL) is a negative co-localisation result and right panel (C-D, *SDC1* gene and TC) is positive co-localisation result. $-\log_{10}$(P) association P-values for phenotype (top: A and C) and $-\log_{10}$(P) association P-values for gene expression (bottom: B and D) at the *FRK* (left: A and B) and *SDC1* locus (right: C and D), 1Mb range.
Abbreviations: LDL: low density lipid; TC: total cholesterol. Adapted from reference: doi:10.1371/journal.pgen.1004383.g001 [78]

Cell and context specificity of regulatory variation

Genetic variations regulate gene expression levels spatially and temporally, thus leading to the differentiation and maintenance of highly specialised cell and tissue types that underlie phenotypic differences across individuals [79]. In general, different cell types are thought to be governed by a set of master regulators – transcription factors and cis-regulatory factors that dictate gene expression for specific cell types [80]. As a result, recent studies in regulatory genomics have been performed across cell types [81]. Due to their accessibility, many eQTL studies have been performed in blood-derived cells, mostly in whole blood and in cell lines, i.e. LCL, a lymphoblastoid cell line that is derived from Epstein Barr Virus (EBV) transformation of

peripheral blood B-lymphocytes. LCL has been very important for the understanding of regulatory variations, particularly as a surrogate for immune cells [68, 82, 83]. However, there remain some questions on the physiological relevance of LCL, since its growth rate and EBV load are both correlated with gene expression levels [84]. Datasets derived from whole blood [75, 85, 86] have also been extensively used. Though informative, there are some drawbacks - whole blood is a 'bulk' tissue, which comprises a variety of cell types. Subtle eQTLs in specific cell types may be diluted when assaying expression across the entire range of cell types [87]. These findings have propelled studies to measure gene expression directly in purified cell types [88] or decompose expression of bulk tissue into specific subsets statistically [89, 90].

Moreover, gene expression levels are also highly dependent on cellular contexts and environmental triggers [87]. A study of purified monocytes derived from 432 European individuals were assessed for gene expression *ex vivo* naïve and stimulated with either interferon-gamma (IFN-r) or lipopolysaccharide (LPS) at 2 and 24 hours (Figure 1-7). The majority of cis-eQTLs were observed only under activation, and many can even possess opposing effect directions [73, 91]. Another study in monocytes found that cis-eQTLs interacted with either age, sex, or smoking status [92] indicating the involvement of environment.

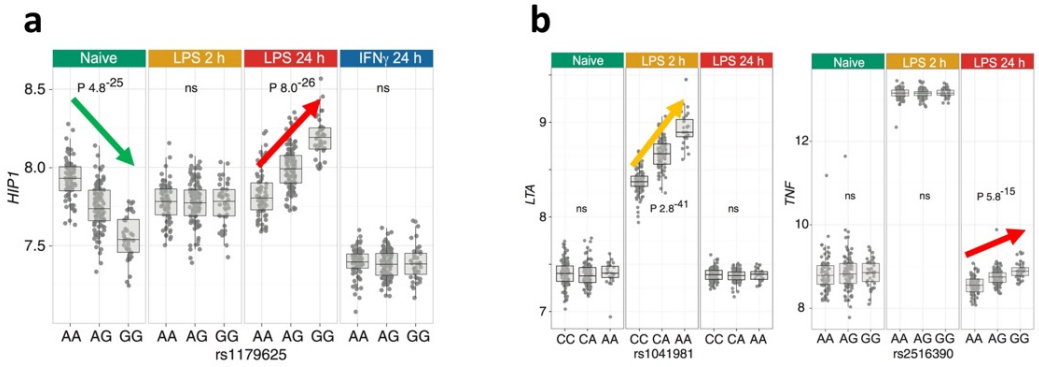


Figure 1- 7**. eQTLs display context and stimulus specificity.**

a) Context-specific directional eQTL effects. Expression of gene *HIP1* in naive and stimulated monocytes stratified on variation of SNP rs1179625. The minor allele [G] decreases expression of *HIP1* in naïve monocytes, but increases expression after 24-hour LPS stimulation. ns, not significant. B) Induction by LPS reveals a stimulus-specific cis-eQTL for *LTA* and *TNF*. Modified from ref [91].

1.3.2 Integrating epigenetics to annotate functional/regulatory variants

An approach that is complementary to eQTL analyses to examine the regulatory function of non-coding genetic variants is to study gene regulation with epigenetics. Epigenetic modifications, a term coined to describe genome-wide chromatin modification, including DNA methylation, histone modifications, chromatin accessibility, microRNA regulations, and 2D chromatin interactions [93], constitute an additional layer of genomic regulation, and may serve as a dynamic link between genotype and phenotype. Such changes in DNA and chromatin structure correlate with changes in chromatin accessibility and transcription factor binding. These epigenomes provide valuable information for prioritizing likely causal variants identified from GWASs. For example, promoters active in a specific cell type can be used to prioritize eQTL variants of that cell type and then be used for identifying potential causal variants.

The Encyclopedia of DNA elements (ENCODE) project (https://www.encodeproject.org/) [94] has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification, and assigns biochemical functions for 80% of the genome, in particular outside of the protein-coding regions. Overall, the project has provided an expansive resource to define functional DNA elements for biomedical research, although the available cell types or cell lines are limited. The cells of closest immune relevance in ENCODE Tier 1 and Tier 2 are LCLs (GM12878), B cells (CD20+) and monocytes (CD14+), as well as T cells (CD4+) and peripheral blood mononuclear cells (PBMC) in Tier 3. A recent ImmunoChip study in Asians [50] took advantage of ENCODE data to map the underlying associated loci. For example, one of the signals (rs73366469) identified in this study was located between two 'general transcription factor' genes, *GTF2I* and *GTF2IRD1*. By integrating the ENCODE data, they found that an indel SNP rs587608058 (r^2 = 0.81), ~1000bp from rs73366469, lay within a conserved enhancer, active chromatin and transcription factor binding sites in LCLs and CD4+ T cells. In addition, this region was found to overlap the transcription start sites for *GTF2I* and *VCF* through chromatin interacting analysis and chromosome conformation capture [95], providing evidence for the potential causal variants and genes at this locus for further study.

The Roadmap epigenomics project (http://www.roadmapepigenomics.org/) [96] integrated analysis of 111 reference human epigenomes to obtain a comprehensive map of the human epigenomic landscape across a large collection of primary cells (including immune cells) and tissues. This map is extremely useful for studies of genome interpretation, gene regulation,

cellular differentiation, genome evolution, genetic variation and human disease. In our meta GWAS analysis of Chinese and European data [54], histone modification markers, including acetylation markers (H3K27ac, H3K9ac) and methylation markers (H3K27me3 and H3K9me3), from blood cell types were used to investigate the potential regulatory function of the target risk loci. For example, there are several genes, including *SRGAP2, SRGAP2D, IKBKE, RASSF5, EIF2D* and *DYRK3*, located within ±200kb of a lead GWAS SNP rs2297550. This GWAS SNP was also found to be a putative eQTL for *IKBKE*, with the SLE risk allele correlated with reduced expression in CD4+ T cells, CD19+ B cells and NK cells, but with increased expression in CD14+ monocytes (Figure 1-5 c). *IKBKE* encodes a noncanonical I-kappa-B kinase (IKK) that is essential in regulating inflammatory responses to viral infection by activating the type I interferon, NF-kB and STAT signalling pathways, suggesting *IKBKE* might be the potential causal gene. Moreover, there is an intense histone acetylation peak around the associated SNP rs2297550, indicating that rs2297550 may be a potential causal variant.

Another recent completed large-scale epigenomic project, the Blueprint project (http://www.blueprint-epigenome.eu/) [97-99], has impressively shown how epigenetic information and analyses can help to study the cellular mechanisms associated with complex human diseases. The Blueprint consortium generated three comprehensive reference panels, including genome (whole genome sequencing), transcriptome (RNA-seq), and epigenome (DNA methylation and histone modification), in three immune cells (Neutrophils, monocytes and T cells) from nearly 200 individuals to characterize the contributions of diverse genomic inputs to transcriptional variation. Summary data from these panels can be accessed through http://blueprint-dev.bioinfo.cnio.es/WP10/ . High-resolution maps of promoter interactions [98] generated by 'Promoter capture Hi-C' (PChi-C) make it possible to study long range regulation in three-dimensional nuclear space. By integrating PCHi-C data with disease-associated SNPs generated by GWAS, we can prioritize the putative target genes for the risk loci. The promoter interactome's map may serve as a more robust method to define cis-eQTL rather than by distance, revealing insights into genomic regulatory mechanisms of diseases.

1.4 From genetics to protein expression

Genetic and epigenetic studies in SLE provide abundant data to address which molecules are likely to have functions in the pathogenesis of the disease. In order to better understand the molecular mechanisms of these genes in the disease, it is enlightening to explore the protein expression of a susceptibility gene. Protein expression is an informative intermediator to minimise the gap between genetics and phenotypes. As proteins are likely to be more directly mediating functional effects on cells than transcripts, one can predict that levels of protein will correlate better with phenotype than mRNA levels. As part of the functional post-GWAS study, I will investigate the protein expression of selected genes that showed robust associations with SLE susceptibility [43], namely Ikaros family members and OX40L, in a range of immune cells, using multi-parameter flow cytometry. In this section, I will summarise what is currently know of their underlying biology and provide justifications for why these genes were chosen for the follow up studies.

1.4.1 Immunophenotyping of PBMC by flow cytometry

SLE is an autoimmune disease involving both the adaptive and innate immune systems. In addition, the target end organs can vary largely across SLE patients. Therefore, the decision of what tissues and cell types are appropriate to perform follow up experiments is of great importance. Fortunately, by integrative analysis of reference human epigenomes from the ENCODE [94] and Roadmap [96] projects with the SLE GWAS findings, I found that most of the regulatory genetic variants are enriched in blood (more details see Chapter 4. The Functional Analysis of SLE Genetics), providing robust evidence for further studies to focus on immune cells in blood. Nevertheless, blood is a bulk tissue that comprises a variety of immune cells, e.g. B cells, NK cells, T cells, monocytes, and etc. Therefore, a comprehensive immunophenotyping to scrutinize the subtle differences across discrete immune subsets is crucial to disentangle the functions of the target molecules in blood.

Immunophenotyping is the analysis of heterogeneous populations of cells for the purpose of identifying the presence and proportions of the various populations of interest based on the types of antigens or markers on the surface of the cells [100]. Cell markers are a very useful way to identify a specific cell population, but they are often expressed on more than one cell type. Flow cytometry has increasingly become a method of choice for the analysis of cellular phenotype and

function in the immune system. It is arguably the most powerful technology available for probing human immune phenotypes, due to its ability to measure multiple parameters on many individual cells in a complex mixture like blood. By evaluating the unique repertoire of cell markers using several antibodies together, each coupled with a different fluorochrome, a given cell population can be identified and quantified. Many immunological cell markers are CD markers and these are commonly used for detection in flow cytometry of specific immune cell populations and subpopulations. In addition, since a wide range of antibody reagents and protocols are available commercially, flow cytometry can be used to access not only the cell-surface protein but also that of intracellular transcription factors. Therefore, immunophenotyping by flow cytometry will be used for the quantification of protein levels of the selected molecules, including the protein products of four individual SLE risk genes: IKZF1, IKZF2, IKZF3 and OX40L. A typical flow cytometry experiment is showed in Figure 1-8.

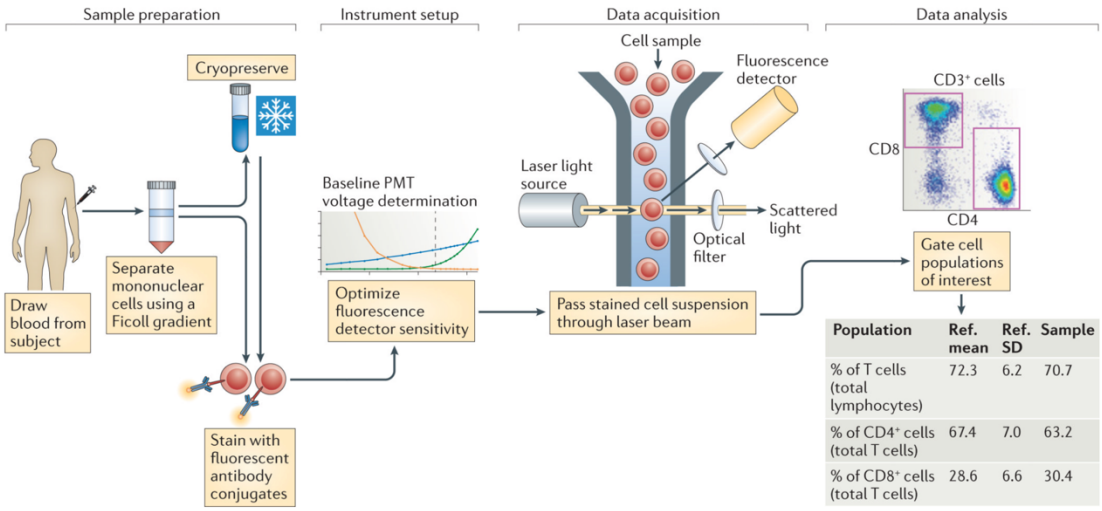


Figure 1- 8**.** A typical flow cytometry experiment

Sample preparation from blood often involves Ficoll gradient separation of peripheral mononuclear cells (PBMC), and sometimes cryopreservation, before staining with the fluorescent antibody conjugates. Instrument setup is to set the voltage gains for the photomultiplier tubes (PMTs) in order to get the optimal detector sensitivity. Data acquisition involves passing the stained cells through a laser beam, and then recording the fluorescence emission from all of the bound antibody conjugates of each single cell by the fluorescence detector. Finally the data analysis reports the cell populations of interest. Ref., reference; SD, standard deviation. Adjusted from reference [100].

1.4.2 The Ikaros family of transcription factors

Immunological homeostasis including proliferation, survival, and cellular differentiation, are mostly governed by transcriptional regulators that mediate appropriate changes in gene expression [101]. The Ikaros gene family encodes zinc finger transcription factors, comprising Ikaros (also called IKZF1), Helios (IKZF2), Aiolos (IKZF3), Eos (IKZF4) and Pegasus (IKZF5), which are involved in lymphoid development and differentiation [102] and are central to successful decision making during haematopoiesis and to the maintenance of a functional immune system. The so-called 'zinc finger' superfamily is characterized by the presence of conserved zinc finger motifs. Differential splicing of exons of the Ikaros gene generates multiple isoforms of several Ikaros family members. These transcription regulators participate in a complex network of interactions with other gene regulatory factors, e.g. activation of STAT4 [103], other family members (e.g. IKZF1 can modulate the expression of IKZF2 [104]) and a raft of other transcriptional regulators modulate their function and regulate important cell-fate decisions during haematopoiesis, particularly in the development of the adaptive immune system [105].

IKZF1, the founder and most studied member of the family, acts as a strong tumour suppressor in B-cell acute lymphoblastic leukemia (ALL) [106]. Deletion of *IKZF1* is of high frequency in BCR-ABL1–positive ALL (a known very-high-risk ALL subtype) and is identified as a strong predictor of poor outcome in ALL patients [107]. Insights from several GWASs, including SLE (Figure 1-9) [108], inflammatory bowel disease (IBD) [109], Stevens-Johnson syndrome (SJS) [110] and Type I diabetes [111], show that the *IKZF1* has been implicated in the susceptibility to autoimmune diseases, indicating the crucial role of *IKZF1* regulation in the maintenance of self-tolerance.

Another two members from the Ikaros family, namely *IKZF2* and *IKZF3*, are shown to be associated at a genome-wide level of significance within the meta-analysis in the largest European GWAS study as well (Figure 1-9) [43]. Helios, encoded by *IKZF2*, is a highly conserved transcription factor expressed in early hematopoietic progenitors of the bone marrow and a high level in thymus-derived regulatory T cells (Treg) [112]. Chromatin-immunoprecipitation (ChIP) experiments indicated that Helios binds to the FoxP3 promoter and knocking-down Helios with siRNA oligonucleotides resulted in down-regulation of FoxP3 [113]. Studies by Alexander T et al found that Foxp3(+) Helios(+) Treg, unlike Foxp3(+) Helios(-) Treg, were significantly increased in SLE patients and expanded in active SLE, suggesting the functional suppressive capacity of Helios in Treg [114].

Aiolos, encoded by *IKZF3*, expresses its highest level in mature peripheral B cells and plays a key role in B-cell differentiation and malignancies [102, 115]. It has also been reported to control T cell death by regulating the expression and localization of Bcl2 [116]. Sun J et al showed that Aiolos-deficient mice develop the symptoms of human SLE, indicating that normal Aiolos function is necessary in the maintenance of immune homeostasis [117]. Moreover, genetic polymorphisms in *IKZF3* have been reported to be associated with the risk of multiple immune-related diseases, not only SLE (Figure 1-8), but also asthma [118], rheumatoid arthritis [119], and ankylosing spondylitis [120] .

1.4.3 The TNF superfamily gene TNFSF4/OX40L

Research over the last decade has shown that one of the most prominent interactions in the TNFR/TNF superfamily members is that between TNFRSF4 (OX40L, CD134) and its partner TNFSF4 (OX40L, CD252) [121]. They are rapidly emerging as key players in their ability to regulate conventional CD4 and CD8 T cells and modulate NKT cell and NK cell function. The primary source of OX40L is likely to be activated antigen presenting cells (APCs), including B cells [122], mature conventional dendritic cells (cDCs) [123], Langerhans cells [124], NK cells, and macrophages [125]. It is noteworthy that OX40L is not expressed on quiescent APCs, but is rapidly up-regulated upon activation. Of interest, OX40L has also been detected on the surface of activated CD4 and CD8 T cells [126]. While the expression of OX40 - the receptor of OX40L, is dependent on the activation of the T cell [127].

GWAS and family-based studies [128] have underlined the importance of TNFSF4 as a SLE susceptibility gene (Figure 1-9), with an increased expression of TNFSF4 transcript in individuals carrying the disease risk allele [129]. The increased expression of TNFSF4 may act by destabilizing peripheral tolerance through inhibiting the generation of IL-10-producing CD4+ type I regulatory T cells [130]. Blocking OX40L has exerted strong reduction in disease activity or a complete lack of disease in multiple animal models of inflammatory and autoimmune disease [121]. The role of TNFSF4 in the pathogenesis of SLE highlights the importance of the role of the T cell-APC interaction in this disease, suggesting this molecule may be an attractive novel therapeutic target. Future studies will be needed to explore the molecular mechanism of OX40L involved in the pathogenesis of SLE.

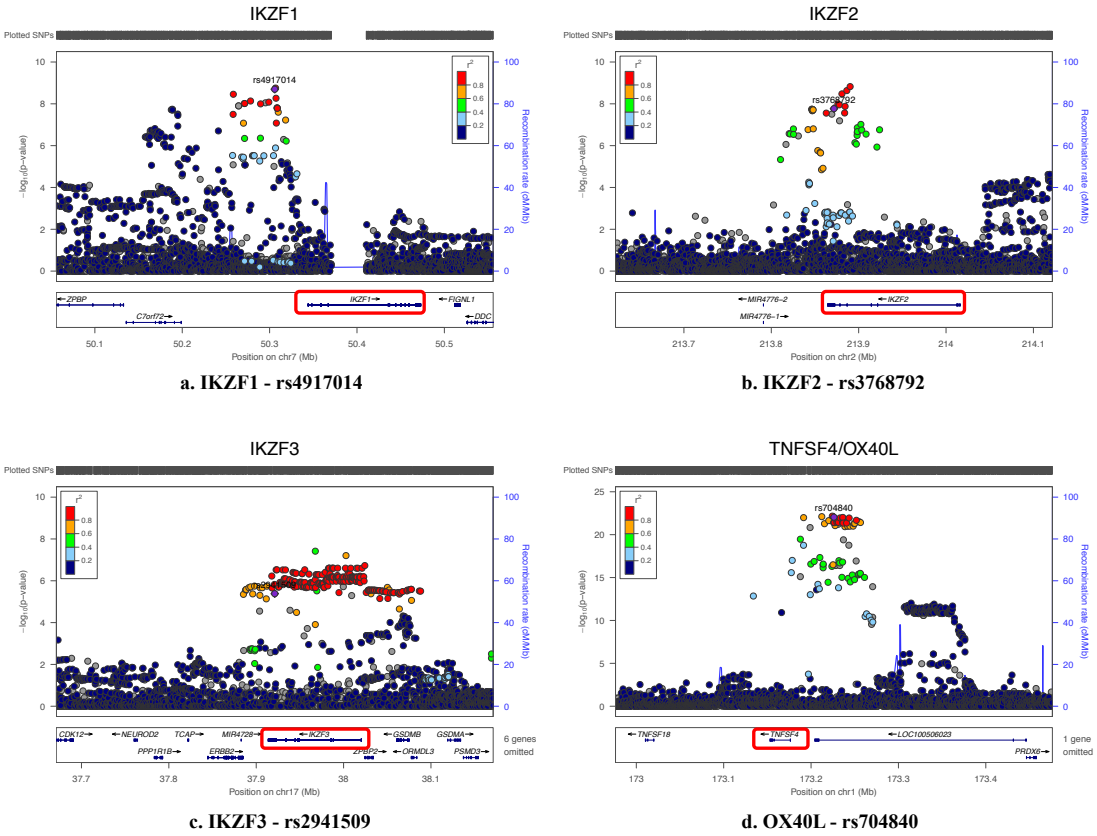


Figure 1- 9**. LocusZoom plots of SLE GWAS signals of target genes.**

These plots were generated using LocusZoom (http://locuszoom.sph.umich.edu/). Data are taken from the Meta-analysis of Chinese and European SLE GWAS [54]. The GWAS SNP (lowest association P-value) at each locus is highlighted. P-values are from a meta-analysis of 3 GWAS cohorts (one European GWAS and 2 Chinese GWASs). Note that the Genome build/LD population is based on gh19/1000 Genome Nov 2014 EUR release.

1.5 Genetic risk loadings and kidney involvement in SLE

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease characterized by a wide spectrum of signs and symptoms varying among affected individuals and can involve many organs and systems, including the skin, joints, kidneys, lungs, central nervous system, and haematopoietic system [1]. Lupus nephritis (LN) is the most common cause of morbidity and mortality. In particular, kidney involvement is one of the most worrisome manifestations of SLE. Patients with kidney disease are likely to have more severe clinical outcomes and a shorter lifespan. 30-60% of adults and up to 70% of children with SLE have renal disease, characterized by the glomerular deposition of immune complexes and an ensuring inflammatory response [131]. Genetic ancestry influences the incidence and prevalence of SLE and kidney involvement, being more frequent in Hispanics, Africans and Asians than in Caucasians [2, 132-134]. Currently, renal disorder in SLE is classified according to the American College of Rheumatology (ACR) renal criteria, i.e., persistent proteinuria and cellular casts (red blood cells, granular, tubular, and/or mixed). However, not all patients will response to the therapy, indicating that additional information focused on mechanism of the tissue involvement is required. Moreover, the early detection of kidney involvement in SLE is useful because early treatment can be applied to reduce the accumulation of renal disability.

Although the exact aetiology of lupus is not fully understood, a strong genetic link has been identified through the application of family [11, 37] and twin studies [10]. SLE does not follow the Mendelian patterns of inheritance, and so it is termed a non-Mendelian disease or complex trait. Complex traits are multi-factorial with both genetic and environmental contributions. In addition, there is no single locus containing alleles that are sufficient for disease in most cases. GWAS has been successfully used to investigate the genetic basis of a disease by comparing a group of patients with a specific disease to a group of healthy controls, which dramatically advanced knowledge of the genetic aetiology of SLE. Our recent review has summarized a total of 84 genetic loci that are implicated as SLE risk [64]. Despite the advances in the genetics of SLE, it is not clear how to utilise genetic information for the prediction of SLE risk or severity in clinical practice. A critical first step is to understand the role of aggregate genetic risk factors, rather than associations of individual alleles with SLE.

A Genetic risk score (GRS) summarizes risk-associated variations across the genome by aggregating information from multiple risk SNPs. There are two basic ways to calculate the GRSs,

i.e. the simplest GRS counts disease-associated alleles and the weighted GRS counts the alleles at each SNP weighted by the log Odds Ratio for each SNP. Because GRSs pool information from multiple SNPs, each individual SNP is not so influential on the summary measurement. Thus, the GRS is more robust to imperfect linkage for any tag SNP and causal SNP, and less sensitive to minor allele frequencies [135-138]. As the number of SNPs included in a GRS grows, the distribution of values approaches normality, even when individual risk alleles are relatively uncommon. Therefore, a GRS can be an effective means of constructing a genome-wide risk measurement that summarises an individual's genetic predisposition to SLE. Hence, we investigated the usefulness of an aggregate measure of risk of SLE renal disease that is based on genetic susceptibility loci identified from SLE GWAS.

Chapter 2. Methods

In this chapter, I will summarise all the methods and materials used in this thesis, including the samples source for the genome-wide association studies and for the functional experiments, the public datasets for the inference of functional context and genes, as well as the software and online tools for the data analyses.

2.1 Samples source

I used two cohorts of unrelated Systemic Lupus Erythematosus (SLE) cases of European descent, including the largest SLE GWAS study (SLE main cohort) [43], and the SLEGEN cohort [41] for the genome-wide association studies. The enrolment of these cases has been described in previous publications [41, 43]. Clinical sub-phenotypes of all patients were documented according to the standard American College of Rheumatology (ACR) classification criteria. Subgroups of patients with kidney involvement, lupus nephritis (LN), or without LN were identified according to the ACR criteria recorded by physicians. To be more specific, patients who had persistent proteinuria defined as a spot urine protein/creatinine ratio of >0.5 or greater than 3+ on a urine test strip, and/or cellular casts including red blood cells (RBC), hemoglobin, granular, tubular or mixed were diagnosed as LN [139].

For the samples in the functional experiments (Section 2.6 Immuno-phenotyping), ethical approval was obtained from the institutional review committee of King's College London (Study Ref: 07/H0718/049). All SLE patients and healthy controls were given information sheets and verbal explanations of what the research entailed. Informed written consent was obtained from all subjects.

2.2 Genome-wide association study (GWAS)

2.2.1 SLE GWAS

SLE GWAS was performed in the SLE main cohort in this thesis. The SLE main cohort comprised 4,036 individuals with SLE (cases) and 6,959 controls, genotyped using either the HumanOmni1-Quad chip or the Illumina HumanOmni2.5 BeadChip [43]. The final SNP set included 644,674

markers that were present on both chips. Four principal components were used as covariates to correct for population structure. For imputation, data were pre-phased using the SHAPEIT algorithm [140] and then imputed to the density of the 1000 Genome reference data (phase 3 integrated set, release 20130502) [74] by my colleagues David Morris and Philip Tombleson (data unpublished). All case-control analysis was carried out using the SNPTEST algorithm [141] by David Morris. In this study, SNPs with imputation INFO scores of <0.7 and MAF < 0.001 were removed from further analysis. After quality control (QC), there are 21,431,070 SNPs left for further analysis. A standard threshold of $P = 5 \times 10^{-8}$ was used to report genome-wide significance. More details about the original GWAS can be found in References [43, 54].

2.2.2 SLE Renal GWAS within SLE cases

Patients with renal disease were considered as more severe than patients without. Hence, I tested genome-wide associations within SLE cases, i.e., associations of patients with renal disease (SLE Renal+) and patients without renal disease (SLE Renal-). The SLE renal GWASs were performed in two cohorts mentioned above, i.e., the SLE main cohort and the SLEGEN cohort. Only patients with clinical renal sub-phenotype were included in this section. Following quality control, the sample size of patients with renal disease (SLE Renal+, more severe) were 1152 and 146 and and patients without renal disease (SLE Renal-, less severe) were 1949 and 378 in the SLE main cohort and SLEGEN cohort, respectively.

Following the same procedures of data processing of the original GWASs, only markers with an IMPUTE info score ≥ 0.7 and MAF ≥ 0.001 were included in further analysis. All case-control analysis was carried out using the SNPTEST algorithm [141]. Moreover, genome-wide association meta-analysis in the SLE main cohort and SLEGEN cohort was performed using R. A standard threshold of $P \leq 5 \times 10^{-8}$ was used to report genome-wide significance and a $P < 1 \times 10^{-5}$ was used to report suggestive associated signals.

All the significant or suggestive SLE/Renal associated loci identified from GWASs were annotated by ANNOVAR [142], a software that utilizes up-to-date information to annotate SNPs, such as inferring their cytogenetic bands, examining their functional consequences on genes, finding variants in conserved regions, or identifying variants reported in dbSNP and the 1000 Genome Project. The required input of ANNOVAR is a list of variants with chromosome, start position, end position, reference nucleotide and observed nucleotides.

2.3 Enrichment of GWAS SNPs in functional epigenomes

2.3.1 GARFIELD enrichment analysis

To systematically characterize the cellular and regulatory contribution of genetic variations implicated in SLE/Renal disease, I used a non-parametric enrichment analysis approach, GARFIELD (GWAS analysis of regulatory or functional information enrichment with LD correction) [143]. Briefly, this method takes summary statistics from genome-wide association studies to calculate the fold of enrichment, i.e., the enrichment of SLE associated variants in functionally annotated regions, at given significance thresholds. The algorithm then tests the statistical significance of the observed enrichment via permutation testing while accounting for LD (linkage disequilibrium), MAF (minor allele frequency) as well as local gene density.

Epigenome regulatory maps

The genome-wide functional regulatory data derived from DNase-seq, i.e., DNase I hyper sensitive sites (DHS), were obtained from ENCODE and the NIH Roadmap Epigenomes [94, 96], which are available in almost all cell types and tissues, providing an overview of the enrichment pattern of the SLE/Renal disease associated variants. DHS data was processed and adapted into the GARFIELD package by the software developer [144].

Processing GWAS summary statistics

To remove possible biases due to linkage equilibrium (LD) between variants, the correlation between two SNPs within a 1MB window is calculated and is considered as independent if a $r^2 \leq 0.1$. The genetic linkage map for the European population was computed from the UK10K sequence data on 3,621 samples from two population cohorts (TwinsUK and ALSPAC) [145] and embedded as part of the GARFIELD package. Next, from the genome wide variants for each phenotype (SLE or Renal disease), the algorithm pruned the SNPs by sequentially removing variants with $r^2>0.1$ and within 1MB window from the most significantly trait-associated variant, and created an independent set of SNPs while retaining the most significant SNPs for each region. Then each independent SNP is annotated and considered as overlapping with a functional element (i.e. DHS) if the SNP itself resides in the according genomic region or at least one of its proxies (within 500kb) in LD with it ($r^2>=0.8$) does.

Calculating fold of enrichment

To see the enrichment of GWAS signals within DHS, I used GARFIELD to calculate the fold of enrichment (FE) as the proportion of the SNPs that fall in DHS as well as have a P-value less than threshold T, divided by the fraction of total number of SNPs within DHS. Specifically,

$$\textit{Fold of Enrichment}_{(t)} = \frac{N_a^t / N^t}{N_a / N}$$

where N is the total number of pruned SNPs, N_a is the total number of SNPs that fall in DHS regions, N^t is the total number of variants with P value less than threshold t, N_a^t denotes the number of SNPs with P value less than t and fall within DHS. That is, the denominator (N_a/N) is the proportion you would randomly expect to fall in DHS regions, while the numerator (N_a^t/N^t) is the proportion you observe at a given GWAS p-value threshold.

Permutation testing and multiple testing adjustment.

To estimate the proportion of times a random selection of SNPs (matching SNPs according to MAF, distance to nearest TSS and number of LD proxies) would have a higher fold of enrichment than observed, permutation testing was applied. I used 10,000 permutations to obtain P-values for fold of enrichment in this thesis. Bonferroni adjustment was applied to account for the multiple-testing burden, which is the effective number (424) of annotations (Supplementary Table 2-1) used for enrichment analysis. The overall procedures of GARFIELD are summarised in Figure 2-1.

In this thesis, fold of enrichment of the SLE GWAS was calculated at eight genome-wide significance thresholds (T) (1×10^{-8} to 1×10^{-1}) in DHS hotspots of all the tissues and for each set tested their enrichment significance at the four most stringent levels (1×10^{-8} to 1×10^{-5}). For the SLE renal GWAS, I calculated fold of enrichment of its statistics at six genome-wide significance thresholds (T) (1×10^{-6} to 1×10^{-1}) in DHS hotspots of all the tissues and for each set tested their enrichment significance at the four levels (1×10^{-6} to 1×10^{-3}).

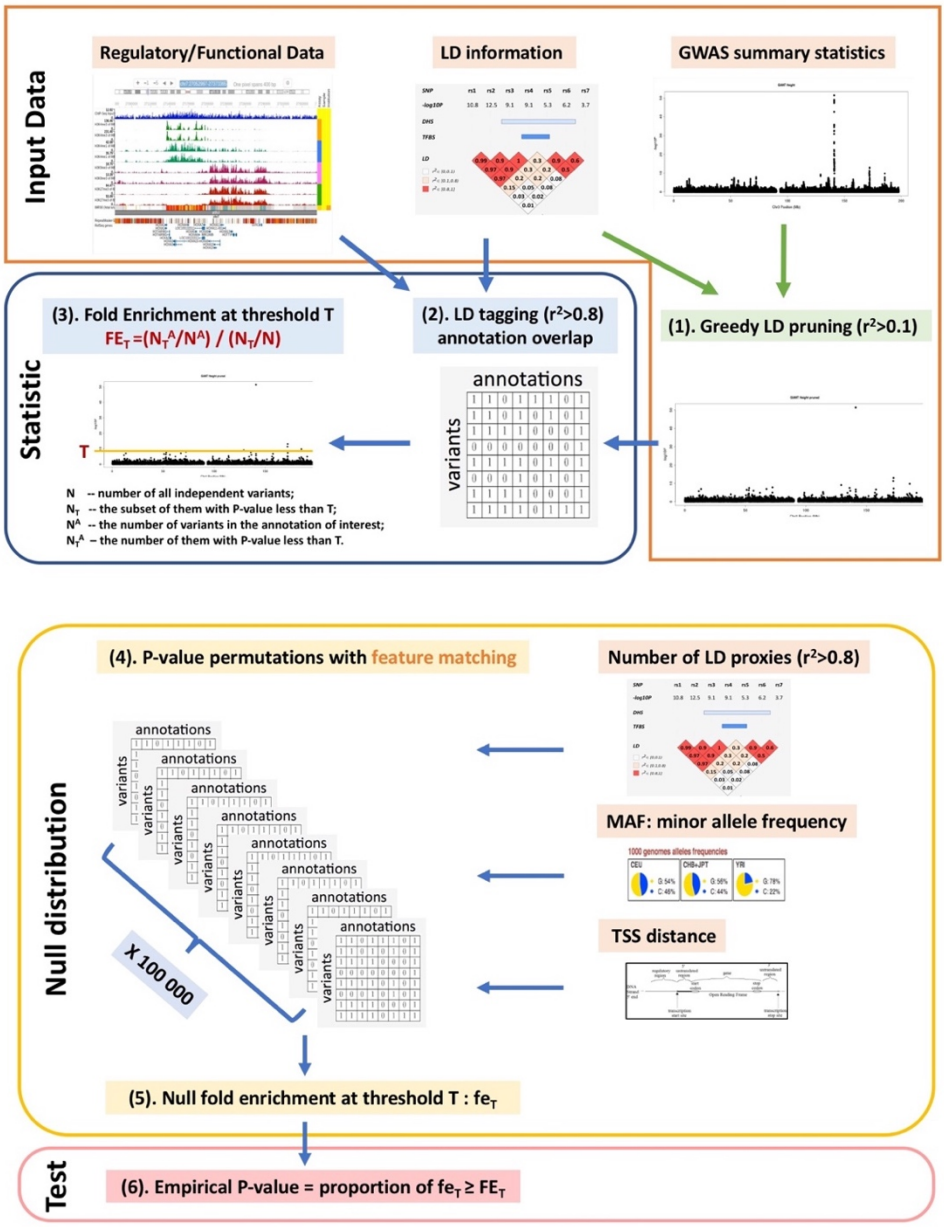


Figure 2- 1**. GARFIELD method flow.**

The 'Input Data' panel consists of genome-wide genomic coordinates for a functional annotation of interest, a list of LD tags for each variant from a reference population, and a set of genome-wide summary statistics (every single SNP's p-value for association with a target trait). Given these inputs, the method performs greedy pruning to reduce the genome-wide variants to an independent set using LD and distance information by removing SNPs with r^2 >0.1 and within a 1Mb window of the most significant SNP. The 'Statistic' panel presents the LD tagging annotation step, which annotates each SNP with a regulatory feature if either the SNP, or a correlated SNP (r^2 >0.8), physically overlaps the feature; and the fold enrichment (FE) statistic at different GWAS P-value thresholds. The 'Null distribution' panel describes the permutation procedure for producing a null distribution for test statistic. Namely, it involves a large number of permutations for each annotation while performing 'feature matching' on variants by MAF, TSS distance and number of LD proxies (r^2 >0.8). Finally, the 'Test' panel shows the calculation of empirical enrichment P-value.

2.3.2 GoShifter – Genomic Annotation Shifter

Though GARFELD shows a comprehensive overview of which annotations the phenotype associated variants are enriched in, specific loci that contribute to the enrichment in particular annotations remain unknown. Therefore, GoShifter [146] is used to explore what phenotype associated loci are enriched in the target annotations. GoShifter is similar in methodology to GARFIELD, but takes different data format as input - a list of target loci / SNPs, which is more flexible for specific analyses and returns more informative output.

Briefly, this algorithm will subset SNPs that are in strong LD (r^2 >= 0.8) with the target SNPs according to 1000 Genome Phase 3 reference genome of European ancestry [74], and randomly shift positions of annotation within each locus whilst fixing positions of SNPs to generate a null distribution. For each annotation (X), it quantifies the observed overlap - the proportion of loci where at least one SNP overlaps. Then it calculates the "delta-overlap" parameter to quantify the effect size of the observed enrichment – the difference between the observed proportion of loci overlapping X and the mean proportion of loci overlapping X under the null. The larger the delta-overlap, the stronger of enrichment. See Figure 2-2 for the schematic of the GoShifter Method.

In this thesis, I downloaded all significant peaks (FDR ≤ 0.01) of DHS for the target cell types from Roadmap Epigenome Project (http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/) and re-formatted using the 200bp region around the peak summit. For each cell type, the proportion of SLE risk SNPs that overlapped with DHS was calculated by GoShifter. It randomly shifts the peaks whilst retaining the positions of the SNPs and re-calculate the frequency of overlap. This was carried out for 1,000 times (permutations) to draw the null distribution. The P-value was calculated as the proportion of iterations for which the number of overlapping loci was equal to or greater than that for the tested SNPs ($P < 0.05$ used as significance cut-off). Finally, I obtained a list of loci that overlap the annotation (informative loci) in the observed data from GoShifter.

Moreover, an ''overlap score'' was calculated to prioritize the informative loci. The overlap score is computed only for the loci that overlap the annotation in the observed data. It is the probability that each locus overlaps an annotation by chance and defined as l_s/n, where l_s is the number of shifting iterations for which at least one SNP within an individual locus overlaps the annotation, and n is the total number of iterations. Loci with low scores are higher-priority candidates for further investigations, as they drive the significant enrichment observations.

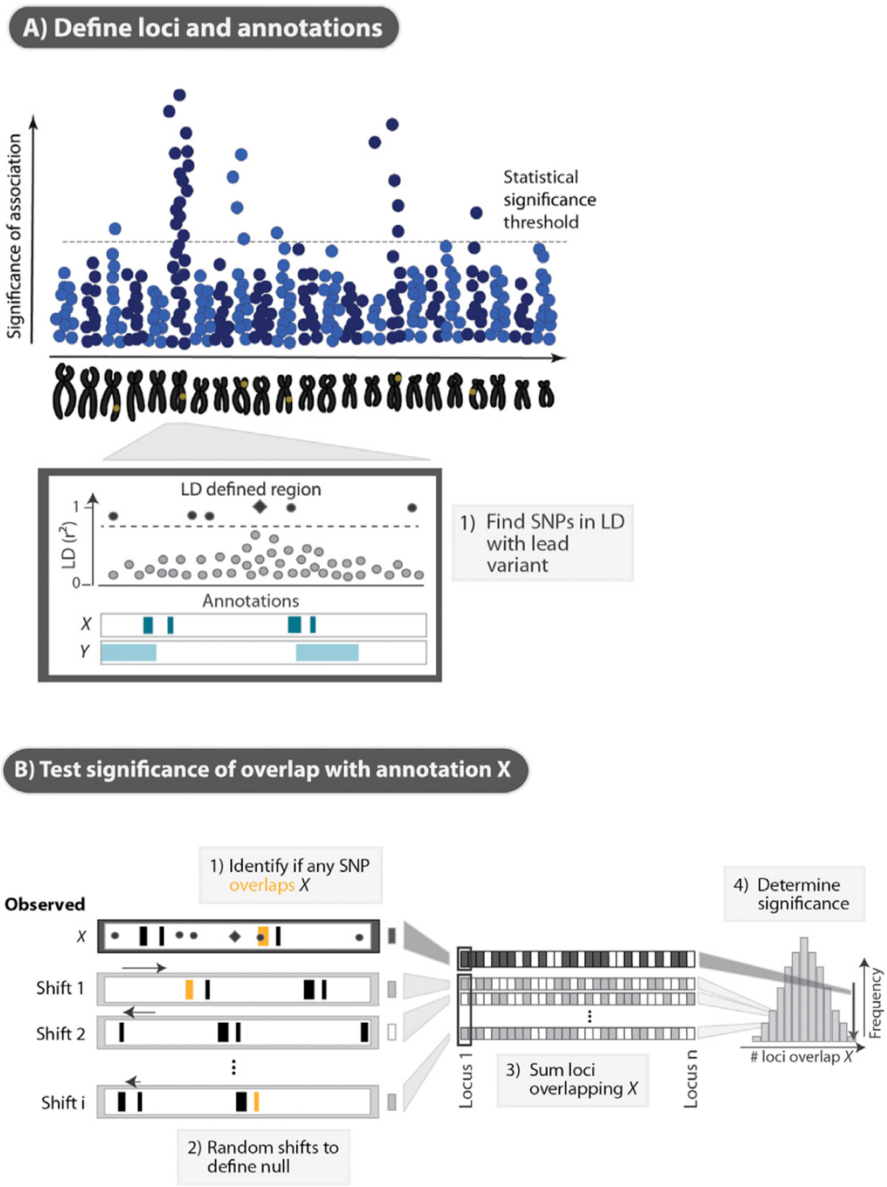


Figure 2- 2**. GoShifter method flow.**

(A) To define the target loci, GoShifter started by searching for SNPs in LD (r2 > 0.8) with the index SNP (i.e. GWAS lead SNP) using the 1000 Genome Project reference data set. (B) To test the statistical significance of an overlap between trait-associated SNPs (square is the index SNP and dot is the SNP in strong LD with the index SNP) and an annotation X (showed in black/yellow lines, of which yellow lines represent the overlapping of target SNP and annotation X). GoShifter first randomly shifted X sites within each locus and quantified the proportion of SNPs overlapping X (shaded boxes). Then it generated a null distribution from these proportions by repeating the shifting process over a large number of iterations. The significance of the observed overlap was estimated by comparing to a null distribution generated by random shifting of X sites (black arrows) within each locus. The P value was computed as the proportion of iterations for which the number of overlapping loci was equal to or greater than that for the tested SNPs. Modified from [146].

2.4 eQTL mapping

2.4.1 Allelic association from SLE GWASs

SLE associated loci taken forward for eQTL analyses in this chapter are derived from a review [64] that summarised all recent SLE GWASs from several ancestries, including European, Asian and Amerindian, as well as subsequent meta-analysis and largescale replication studies. In this chapter, I defined the SLE 'lead GWAS SNP' at each locus according to their publications, either the SNP with the lowest P-value or the SNP with the greatest evidence to be a missense variant. I omitted SNPs located within the expanded Major Histocompatibility Complex (MHC) region (chr6: 29,000,000 – 35,000,000) and non-autosomal regions. In total, 95 GWAS SNPs, summarised in Supplementary Table 2-2, were taken for further eQTL analyses.

2.4.2 eQTL cohorts

Genome-wide gene expression and genotyping data sets applied in this thesis to eQTL analysis are summarized in Table 2-1. Data sets in ex vivo B cells [73], NK cells (unpublished, from Professor Julian Knight's group) and monocytes (naïve, and stimulated by LPS and IFN) [91] are derived from European population. Data sets in CD4+ T cells and CD14+ monocytes of European American (EUR), African American (AFR) and East Asian (EAS) populations are from Towfique Raj et al [147]. A Whole blood data set of 2,765 individuals is from the Consortium for the Architecture of Gene Expression (CAGE) project, which consisted of data from five cohorts [148]. A data set with LCL is part of the HapMap 3 Project, which included genotypes and gene expression from five populations: African (AFR), American (AMR), East Asian(EAS), European (EUR), and South Asian (SAS) [149].

Table 2- 1. Cohorts for eQTL analyses

Study	Platform	Ancestry [a]	Cell type [b]	Sample size
Fairfax et al.[73, 91]	Illumina Human HT-12 v4 BeadChip	EUR	B cell	280
			Monocyte (Naïve)	432
			Monocyte (LPS2)	261
			Monocyte (LPS24)	322
			Monocyte (IFN)	367
			NK cell (not published)	245
Raj et al. [147]	Affymetrix GeneChip Human Gene ST 1.0 microarrays	EUR	Monocyte	211
			CD4+ T cell	213
		AFR	Monocyte	112
			CD4+ T cell	112
		EAS	Monocyte	78
			CD4+ T cell	82
HapMap3 project [149]	Sentrix Human-6 Expression BeadChip v2	AFR	LCL	159
		AMR		42
		EAS		158
		EUR		86
		SAS		75
CAGE [148]	Illumina Whole Genome Expression BeadChips (HT12 v.3, HT12 v.4)	EUR	Whole blood	2765

a. Abbreviations for all the populations/ancestries are: EUR = European; AFR = African; EAS = East Asian; AMR = American; SAS = South Asian.

b. Monocyte (LPS2): gene expression levels were detected after exposing primary monocytes to lipopolysaccharide (LPS) for 2 hours of LPS stimulated. Similarly, LPS24 referred to 24 hours of LPS stimulation. Monocyte (IFN): gene expression level was detected after primary monocytes were exposed to interferon-γ (IFN) for 24 hours.

2.4.3 MatrixEQTL

All eQTL analyses assumed an additive model and were performed using linear regression, as implemented in the R package, MatrixEQTL [150]. *Cis* associations were limited to the variants and probes that are less than 1Mb apart (i.e. a 2Mb window centred on a target position), while

trans associations were tested for all SNP-probes elsewhere. For both cis and trans analyses, I defined a False Discovery Rate (FDR) of 0.01 by including tests across all loci for genes within a ±1Mb window, as calculated by MatrixEQTL, which supports separate p-value thresholds and FDR calculation for cis- and trans-eQTLs. Figure 2-3 shows the general work flow of MatrixEQTL analysis applied in this thesis.

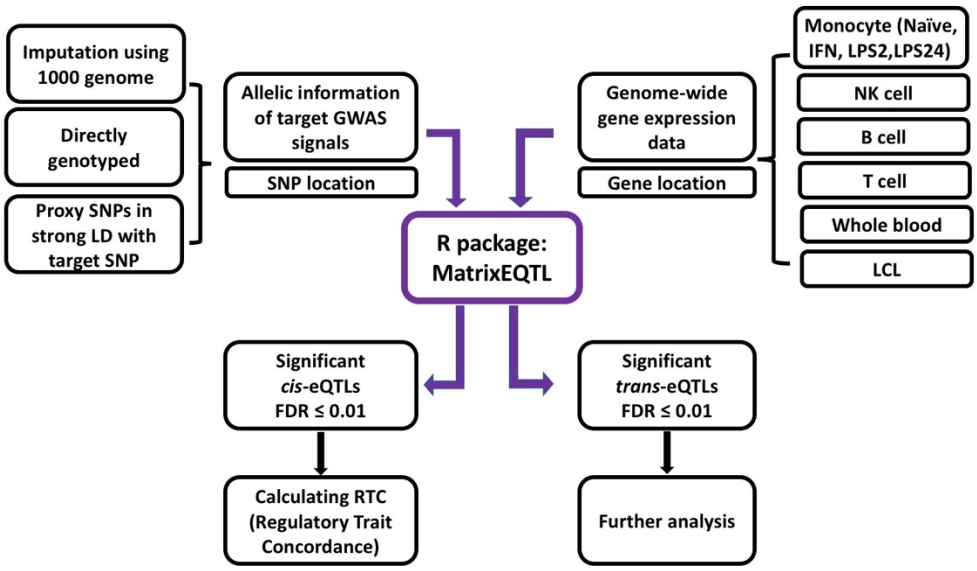


Figure 2- 3. General work flow of MatrixEQTL.
Genotyping data of SLE associated SNPs and gene expression data from data sets shown in Table 3-1 are extracted and formatted as input file for R package MatrixEQTL. SNP-gene pairs with a FDR ≤ 0.01 were exported for further analyses.

2.4.4 Co-localization analysis – Regulatory Trait Concordance (RTC) score

Interrogating the GWAS SNPs for significant association with gene expression has been widely employed to explain GWAS results [151, 152]. However, the ubiquity of regulatory variation throughout the human genome [153] makes it very likely that the overlaps between the eQTLs and the complex trait loci are coincidental. This is due to the linkage disequilibrium (LD) of the genome, which makes functionally uncorrelated variants statistically correlated. This will become particularly problematic as statistical power increase in eQTL with the growing sample size of the study cohort, as well as the availability of summary eQTL results from the public data browser [75, 76]. In these cases, many observed eQTL and GWAS associations are likely to be driven by

two independent causal variants. Therefore, overlaps between eQTL and GWAS loci will need to be evaluated in the context of the genome's correlation structure.

In this thesis, an empirical methodology (regulatory trait concordance – RTC) developed by Nica et al [77] is implemented. I calculated the RTC scores [77] for all eSNP-eGene pairs with FDR $\leqslant$ 0.01 to test whether observed associations between SNPs and expression levels of local genes were due to chance. This score is generated by testing the null hypothesis that the GWAS-associated SNP and the best eSNP (within an interval between two recombination hotspots) are tagging two separate effects and the observed eQTL is purely due to the LD between the GWAS-associated SNP and the 'true' eSNP, that is, to distinguish the accidental co-localisations from the true sharing of causal variant. The process of calculating the RTC score for each eSNP-eGene pair is as demonstrated in Figure 2-4.

Firstly, genotypes for SNPs locating in between a recombination hotspot [154] for a target region as well as expression data of the target eGenes are subset from available data sets. Then I perform a linear regression on gene expression against genotypes and obtain the residuals from regression, as well as the Top eSNP. Following a second round of linear regression of residuals against the best eSNP, I rank all the SNPs in the interval and calculate the RTC using the following formula:

$$\boldsymbol{RTC} = \frac{\boldsymbol{N_{snps}} - \boldsymbol{Rank_{gwas\ snp}}}{\boldsymbol{N_{snps}}}$$

Where $\boldsymbol{N_{snps}}$ is the total number of tested SNPs in between the recombination hotspot region; $\boldsymbol{Rank_{gwas\ snp}}$ is the rank of the lead GWAS SNP according to the *P*-value from the second round of linear regression, i.e., $\boldsymbol{Rank_{gwas\ snp}}$ = 0 if the GWAS SNP is the same as the eSNP, that is, the GWAS SNP has the largest impact on the eGene. Given this, the RTC score will always be in the range of 0 < RTC ≤ 1, with values close to 1 indicating the co-localisation of the GWAS effect and the eQTL effect. In this thesis, I used a cut-off of RTC ≥ 0.95 as significant co-localisation of these two traits for causality inference. As RTC scores have a uniform distribution, setting an RTC threshold of 0.95 sets the type I error rate to be 0.05. We suggest that the gene expression analyses provide some support for likely causal genes, but we note that proof of true causality through altered gene expression will only be achieved by additional experimentation.

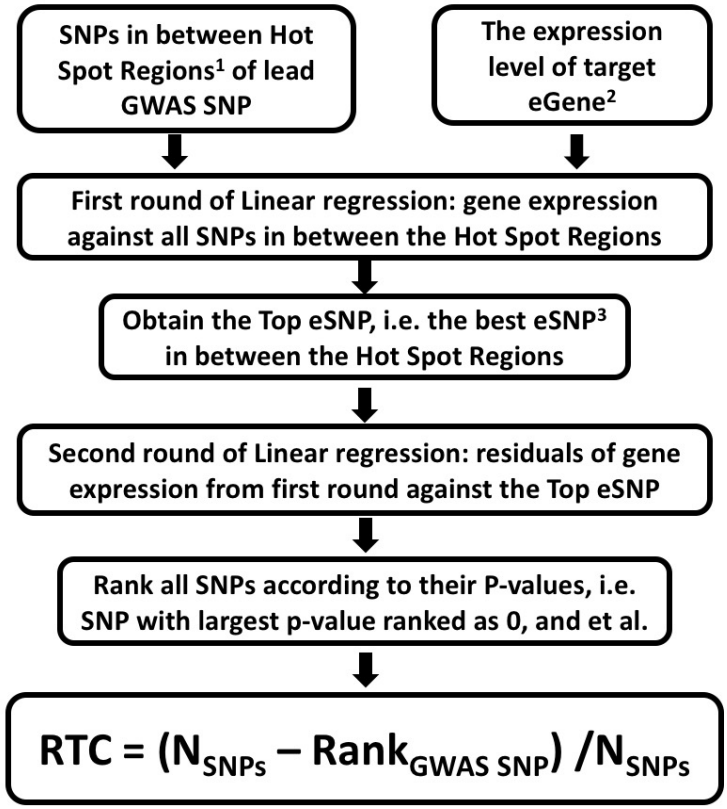


Figure 2- 4. General procedures of calculating RTC scores.

[1]Hot Spot Regions refer to regions in a genome that exhibit elevated rates of recombination relative to a neutral expectation [154]. [2] eGene refers to a gene whose expression level is correlated with genomic variants. [3]eSNP refers to a SNP that is correlated with the expression of one or more genes.

2.5 Genetic risk loadings

2.5.1 GRS calculation

Genetic risk scores (GRS) were calculated using R according to the method described by Hughes et al [155], taking the number of risk alleles (i.e., 0, 1 or 2) for a given SNP and multiplying it by its corresponding β coefficient, i.e. the natural log of its odds ratio (OR). The cumulative risk score in each subject was calculated by summing the risk scores from the loci summarized in the recent review [64] for calculating the cumulative genetic risk score:

$$\text{Genetic risk score} = \sum_{i}^{n} G_i \beta_i$$

where n represents the number of SLE risk loci, G_i is the number of risk alleles at a given SNP, and β_i is indicates the effect size of the risk SNP i.

In this study, an overall GRS was created using 78 SLE susceptibility loci (without the X chromosome), consisting of 95 SNPs in the non-MHC region and 7 independent SNPs in the MHC region (defined as chr6: 29,000,000 – 35,000,000). The allele information (i.e., risk allele, MAF and OR) for the 95 non-MHC SNPs are summarized from the published SLE GWAS [64], while for the MHC region, the allele information were derived from the SLE main cohort [43] by step-wise conditional analysis, see Table 2-2. Each GRS for both SLE cohorts [41, 43, 46] was generated by weighting each risk allele using the effect size reported by their corresponding studies as shown in Supplementary Table 2-2 and Table 2-2. Analysis was carried out using R 3.3.2 (https://www.r-project.org/).

Table 2- 2. Independent SNPs in the MHC region.

SNP	chr	pos(hg19)	Risk allele	Alt allele*	OR	P value in SLE main cohort
rs115032094	6	31340001	G	A	1.70	6.50E-08
rs146903072	6	31847180	G	A	2.49	5.00E-14
rs2596504	6	31319029	T	C	1.97	3.68E-75
rs3134978	6	32651477	T	A	2.32	5.34E-99
rs389884	6	31940897	G	A	2.54	2.70E-102
rs9270782	6	32569119	G	A	1.22	6.48E-11
rs9271636	6	32592119	G	C	1.24	9.43E-11

*Alt allele: the alterative allele of SNP.

2.5.2 GRS categories

Since a continuous score is difficult to interpret on an individual level when a physician needs to explain the results of the GRS to a patient, we partitioned SLE patients into different categories of renal risk. These categories were created from the mean and SD from the control samples (SLE renal-) that were specific to each population. The seven groups were defined as 0.25, 0.75 and 1.25 SD from the mean; the extreme groups were less than 1.25 or greater than 1.25 SD from the mean. To be more detailed, the ranges for the seven groups are defined as: ($-\infty$, mean - 1.25SD), (mean - 1.25SD, mean – 0.75SD), (mean – 0.75SD, mean - 0.25SD), (mean - 0.25SD, mean + 0.25SD), (mean + 0.25SD, mean + 0.75SD), (mean + 0.75SD, mean + 1.25SD), (mean + 1.25SD, $+\infty$). The division of our score into seven groups provided a robust distribution that enabled us to break down the groups with the highest and lowest risk but still ensure that there were statistically sufficient numbers of individuals with renal disease and without in the

extreme categories. To avoid exaggerating the odds ratio associated with a specified subject group, I used the largest subset of individuals that contained the mean of SLE patients without renal disease as the reference group [136]. These individuals can be regarded as those with the mean risk of the assessed population of SLE cases. Within each dataset, I used a chi-square test to study the association of *allGRS* with renal disease and compared each group of *allGRS* with the reference group 4. The chi-square test for trend was used to calculate p-values for trend. We also calculated the odds ratios of renal disease comparing the individuals in the highest *allGRS* group (G7) with those in the lowest *allGRS* group (G1).

2.5.3 Receiver Operator Characteristic (ROC) curves for model evaluation

To determine how well the *allGRS* discriminates cases (SLE renal+) and controls (SLE renal-), I generated receiver operator characteristic (ROC) curves by plotting the sensitivity for the continuous *allGRS* score against '1- specificity' and calculating the area under the ROC curve (AUC) for each cohort. To assess the degree to which the age of SLE onset contributes to the prediction, I plotted ROC curves for *allGRS* that excluded and included the age of disease onset. The AUC was compared by use of a non-parametric approach, as described by De Long and colleagues [156]. When available, additional covariates were considered: in the SLE main cohort, sex and four principle components from its SLE GWAS were added to the model; and in the SLEGEN cohort, all the participants are women, thus only the three principle components from the original GWAS were added to the model.

2.6 Immuno-phenotyping by Flow Cytometry

2.6.1 PBMC isolation

This work is approved by the institutional review committee of King's College London **(Study Ref: 07/H0718/049)**. 20 patients with SLE were recruited from Kidney Clinics of Guy's Hospital, King's College London. The diagnosis of SLE was established according to the 1982 revised American College of Rheumatology criteria. Disease activity was evaluated with the SLE disease activity index score (SLEDAI) [157] by experienced clinicians. Clinical characteristics of patients with SLE are summarised in Table 2-3. 22 healthy individuals matched for gender and age were recruited as controls. All the blood samples are collected after obtaining informed consent.

Peripheral blood mononuclear cells (PBMCs) were isolated from heparinized peripheral venous blood by using Histopaque gradients (Sigma, #H8889) and Leucosep™ Centrifuge Tubes (VWR, #227290). PBMCs were washed twice using phosphate-buffered saline (PBS) (Life Technologies, #14190169) with 2% Bovine Serum Albumin Powder (BSA) (SLS, #A3912) and re-suspended in PBS.

Table 2- 3. Clinical characteristics of SLE patients.

Patient	AgeOnset	SLEDAI	SEX	Malar	Photosensitive	Oral ulcers	Arthritis	Serotosis	Renal	Anti-dsDNA	C3_low	C4_low
P1	27	4	2	2	2	2	1	2	na	na	na	na
P4	32	na	1	1	1	1	1	2	1	1	2	2
P6	39	na	2	1	1	1	2	2	na	na	na	na
P7	31	12	2	1	1	1	2	2	2	na	na	na
P8	38	6	2	1	2	2	1	2	2	na	na	na
P10	na	2	2	2	2	1	2	na	na	na	na	na
P11	23	10	2	1	1	1	1	1	1	na	na	na
P13	na	8	2	na	na	na	na	na	na	na	na	na
P14	40	4	2	1	1	2	1	na	na	na	na	na
P15	na	2	2	2	2	1	1	na	na	na	na	na
P16	23	8	2	1	1	2	1	2	1	1	1	1
P17	25	4	2	2	1	1	1	1	2	1	2	2
P18	25	6	2	2	2	1	1	1	2	2	2	2
P19	39	na	2	1	1	2	1	2	2	2	2	2
P20	26	4	2	2	2	2	1	2	2	2	2	2
P21	21	2	2	na	na	na	na	na	na	na	na	na
P22	na	na	1	1	2	1	1	1	na	na	na	na
P23	16	2	2	1	1	1	1	2	2	1	1	1
P24	13	8	2	1	2	2	2	2	1	na	na	na
P26	na	na	2	na	na	na	na	na	na	na	na	na

Notes: Sex: 1=[male], 2=[female]; Others: 1=[YES], 2=[NO]; na: not available. SLEDAI was recorded at the time of sample collection while other clinical information was extracted from historical databases.

2.6.2 Flow cytometry for Immuno-phenotyping

The experiment was designed to measure the protein level of OX40L, IKZF1, IKZF2, and IKZF3 across various cell types simultaneously. As showed in Table 2-4, each sample of PBMCs was divided into 16 proportions and stained with the appropriate antibody-fluorescence conjugates both on the surface and inside the cells. PBMCs were stained using the best concentration of each antibody obtained by gradient titration, summarized in Table 2-5.

The detailed protocol of PBMC surface staining to determine the immune sub-populations is showed in Appendix C. After staining, PBMCs were washed and re-suspended in staining buffer (BD Horizon™, #563794), and fixed in fixative buffer (BD Cytofix™, #554655). Intracellular staining of IKZF1, IKZF2, and IKZF3 or isotype control were performed according to the manufacturer's protocol. The stained cells were analysed within 24 hours in a BD Fortessa System (BD Biosciences, San Jose, CA), followed by analysis using FlowJo software (Tree Star, San Carlos, CA). Figure 2-5 outlines the work flow of this experiment. The gating strategies for

the immune subsets in this experiment mainly follow the standardized Human Immunophenotyping Consortium (HIPC) phenotyping panel [100] outlined in Figure 2-6, with minor adjustments.

For the data presentation and statistical analysis, I used R (**http://www.r-project.org/**) and GraphPad Prism 7 (**http://www.graphpad.com/**). Generally, quantitative data were expressed as the mean with the 95% confidence interval (95% CI). Statistical significance was determined by analysis of variance or Student's t-test. Correlation tests were determined by Spearman rank correlation or Pearson correlation when appropriate. All tests were followed by adjusting for multiple comparisons using Benjamini & Hochberg (1995), also known as the FDR (false discovery rate) [158]. A FDR (q-value) ≤ 0.05 was considered as statistically significant.

Note: For the FACS experiment, I worked closely with my colleague Mr. Christopher Pinder at the beginning (October 2014 ~ November 2014). Mr. Pinder designed the experiment and showed me how to perform the experiments. I then took over this project after he left for a new job in Imperial College London (November 2014 ~ now).

Table 2- 4. Experimental design for measuring the protein expression of IKZF1-3 and OX40L.

Panels	IKZF IC	IKZF1	IKZF2	IKZF3
Panel 1	**Isotype control FITC** OX40L PE	IKZF1 FITC **Isotype control PE**	IKZF2 FITC OX40L PE	IKZF3 FITC OX40L PE
Panel 2	**Isotype control FITC** OX40L PE	IKZF1 FITC **Isotype control PE**	IKZF2 FITC OX40L PE	IKZF3 FITC OX40L PE
Panel 3	**Isotype control FITC** OX40L PE	IKZF1 FITC **Isotype control PE**	IKZF2 FITC OX40L PE	IKZF3 FITC OX40L PE
Panel 4	**Isotype control FITC** OX40L PE	IKZF1 FITC **Isotype control PE**	IKZF2 FITC OX40L PE	IKZF3 FITC OX40L PE

Note that due to the markers availability and the multi-colour limitation of the flow machine, IKZF1/2/3 were all stained with secondary fluorescence, i.e., Anti-IKZF1 IgG + IgG FITC. It will then generate 4 samples for each panel including the isotype control. The OX40L was stained on the surface with anti-OX40L PE, and the PE isotype control were adding to the samples other than the ones with FITC isotype control.

Table 2- 5. Summary of full panels' fluorescent antibody conjugates.

Panel	Marker	Fluorophore	Brightness [a]	Provider	Catalogue	Best Concentration (µL in 50 µL master mix)
All	IKZF1	NA	NA	R&D Systems	CBGA0212111	1.0
	IKZF2	NA	NA	Santa Cruz Biotechnology	sc-9864	1.0
	IKZF3	NA	NA	Santa Cruz Biotechnology	sc-160013	1.0
	Normal Goat IgG Isotype control	NA	NA	Santa Cruz Biotechnology	sc-2028	0.5
	IKZF Secondary	Alexa Fluor 488	3	BD Biosciences	557721	0.5
	OX40L	PE	5	Cambridge Biosciences	326308	5.0
	Mouse IgG1 Isotype control	PE	5	Cambridge Biosciences	400114	5.0
	Viability Dye	Live/Dead Blue	NA	Life Technologies	L-23105	1.0
Panel 1 (T-cells)	CCR7	Brilliant Violet 421	5	BD Biosciences	562555	5.0
	CD4	PE/Texas Red	4	Life Technologies	MHCD0417	4.0
	CD45RA	Brilliant Violet 785	3	Cambridge Biosciences	304140	4.0
	CD38	APC	5	Cambridge Biosciences	356606	2.0
	CD8	Alexa Fluor 700	2	Cambridge Biosciences	300920	4.0
	CD3	APC/Cy7	2	Cambridge Biosciences	300318	4.0
	HLA-DR	PerCP/Cy5.5	3	Cambridge Biosciences	307630	4.0
	CXCR3	Brilliant Violet 605	5	Cambridge Biosciences	353728	5.0
	CCR6	PE/Cy7	4	Cambridge Biosciences	353418	2.0
Panel 2 (Treg Cells)	CD25	Brilliant Violet 605	5	Cambridge Biosciences	302632	20.0
	CD4	PE/Texas Red	4	Life Technologies	MHCD0417	4.0
	CCR4	APC	5	Cambridge Biosciences	359408	1.0
	CD127	PE/Cy7	4	Cambridge Biosciences	351320	8.0
	CD45RO	Brilliant Violet 785	3	Cambridge Biosciences	304234	8.0
	CD3	APC/Cy7	2	Cambridge Biosciences	300318	4.0
	HLA-DR	PerCP/Cy5.5	3	Cambridge Biosciences	307630	4.0
Panel 3 (B-cells)	CD24	PerCP/Cy5.5	3	Cambridge Biosciences	311116	12.0
	CD19	Brilliant Violet 421	5	Cambridge Biosciences	302234	4.0
	CD27	Brilliant Violet 785	3	Cambridge Biosciences	302832	8.0
	CD38	APC	5	Cambridge Biosciences	356606	2.0
	CD20	PE/Cy7	4	Cambridge Biosciences	302312	4.0
	CD3	APC/Cy7	2	Cambridge Biosciences	300318	4.0
	CD138	Alexa Fluor 700	2	Cambridge Biosciences	356512	20.0
	IgD	Brilliant Violet 605	5	BD Biosciences	563313	20.0
Panel 4 (NK cells, Dendritic cells, Monocytes)	CD56	Brilliant Violet 605	5	Cambridge Biosciences	318334	8.0
	CD123	APC	5	Cambridge Biosciences	306012	4.0
	CD11c	PE/CF594	4	BD Biosciences	562393	4.0
	CD16	Brilliant Violet 785	3	Cambridge Biosciences	302045	4.0
	CD3	APC/Cy7	2	Cambridge Biosciences	300318	4.0
	CD19	Brilliant Violet 421	5	Cambridge Biosciences	302234	4.0
	CD20	PE/Cy7	4	Cambridge Biosciences	302312	4.0
	CD14	Alexa Fluor 700	2	Cambridge Biosciences	301822	4.0
	HLA-DR	PerCP/Cy5.5	3	Cambridge Biosciences	307630	4.0

a. Brightness: The Fluorophore Brightness Index Score is a relative indication of fluorescence intensity above the background for each fluorophore antibody conjugation (1=dim, 5=brightest). NA, not available.

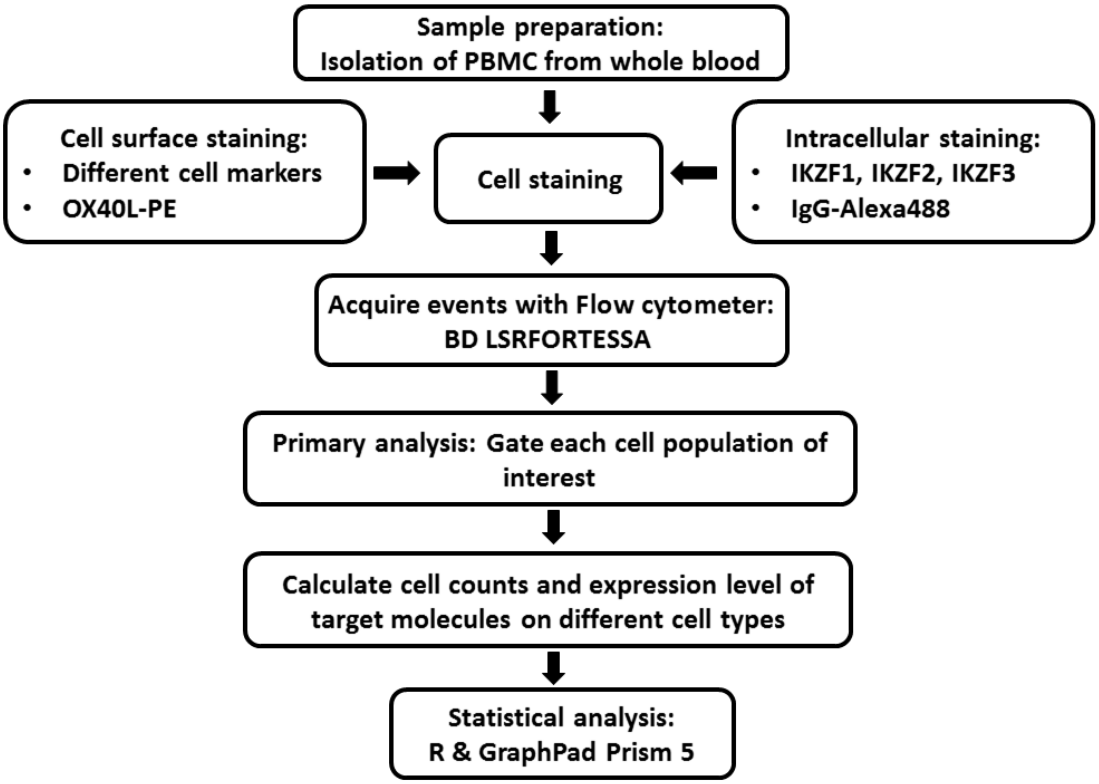


Figure 2- 5**.** Work flow of flow cytometry.

Draw blood from volunteers (SLE patients and healthy controls) and separate PBMC using a Histopaque. Stain cells with fluorescent antibody conjugates, including surface staining for cell markers and OX40L, as well as intracellular staining for IKZF family and their corresponding secondary antibodies. After the staining, setup the instrument (BD LSRFORTESSA) to acquire and record events. Perform primary data analysis using FlowJo_V10 (**http://www.flowjo.com/**) to gate the target populations and export the numerical data, comprising cell counts of all the subpopulations being analysed and the protein expression level of IKZF1, IZKF2, IKZF3 and OX40L on these populations. Statistical analysis and data presentation are carried out by R (**http://www.r-project.org/**) and GraphPad Prism 7 (**http://www.graphpad.com/**).

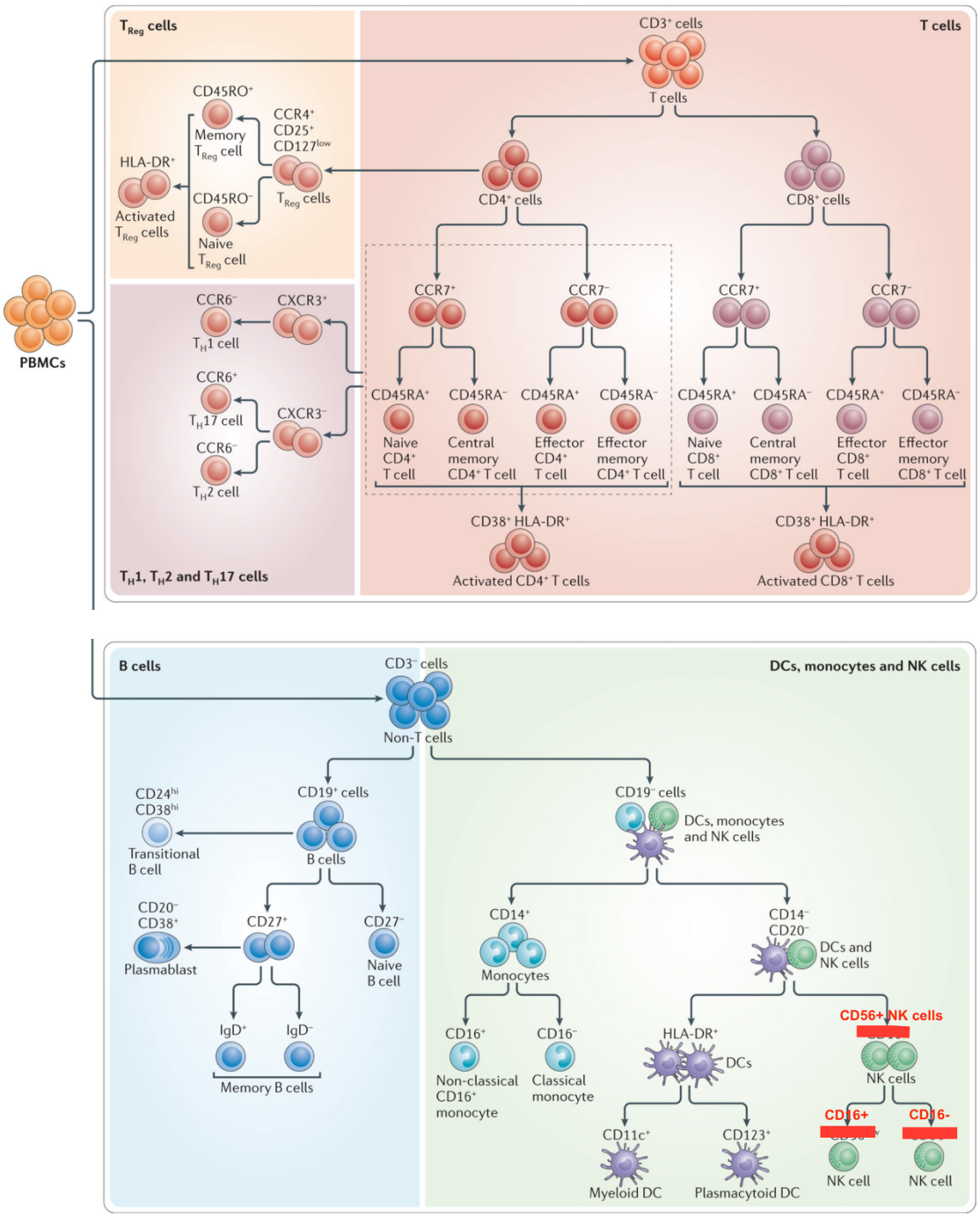


Figure 2- 6. **Identification of immune cell subsets by eight-colour antibody staining.** The figure shows the cell populations that can be identified using the markers targeted by each of the five antibody cocktails of the Human Immunophenotyping Consortium (HIPC) phenotyping panel shown in TABLE 5-2. CCR, CC-chemokine receptor; CXCR3, CXC-chemokine receptor 3; DC, dendritic cell; NK, natural killer; PBMC, peripheral blood mononuclear cell; Th, T helper; Treg, regulatory T. The classification of NK cells were revised according to the corrigendum of the original article [100].

Chapter 3. Expression Quantitative Trait Loci (eQTL) Mapping and Causality Inference

Chapter summary

Recent GWASs have identified more than 80 loci underlying susceptibility to SLE. However, 85~90% of these loci are located outside of protein-coding genes, suggesting that the mechanism underlying their role in disease risk involves changes in gene regulation; they likely exert their functional effect through altering gene expression rather than by altering protein structure [62, 159]. Thus, it is of particular importance to elucidate the actual functions of variants identified as associated with disease. In the post-GWAS era, the greatest challenge is in combining GWAS findings with additional molecular data to functionally characterize the relationship between genetic variants and complex traits. It is now possible to investigate the effect of risk variants on intermediate molecular levels, for example, gene expression, methylation, protein abundance or metabolite levels, due to the advances in a variety of ~omics techniques. In this Chapter, I will focus on studying gene expression (transcription) as an intermediator to illustrate the potential links between genotypes and phenotype.

I performed eQTL (expression Quantitatively Trait Loci) mapping, incorporating co-localization analysis of genetic association, in immune cells (i.e., T cells, B cells, Monocytes, NK cells, as well as whole blood), in order to identify likely causal genes. I found that eQTLs present in a diverse set of immune cells, encompassing both the innate and adaptive immune responses for almost half of the loci. Through comprehensive eQTL mapping using a variety of available datasets, I have identified cell-specific and context-specific eQTLs. By mapping eQTL across populations in LCL, I demonstrated the effectiveness of trans-ethnic eQTL mapping to identify potential causal genes. Furthermore, I have identified trans-eQTLs, providing evidence for further functional studies to elucidate the underlying regulatory networks implicated in the pathogenesis of SLE.

3.1 Discovery of SLE candidate-causal eQTLs (eSNPs) and eGenes

In order to highlight potential causal genes at the susceptibility loci, the 95 SLE risk variants at 77 loci summarized in my review [64] (Supplementary Table 2-2) were tested for correlation with

gene expression in ex vivo in various cell types, including T cells, B cells, natural killer (NK) cells, stimulated and resting monocytes, whole blood and LCL (EBV-transformed lymphoblastoid cell lines), and across populations in some of the cell types (i.e., CD4+ T cells, CD14+ Monocytes, and LCL), comprising cells from Europeans, East Asians, and African Americans. Details of the eQTL cohorts can be found in Chapter 2, Section 2.4.2. In addition, I calculated RTC scores to test if the overlap of eQTL and disease association is coincidental. A RTC score threshold of 0.95 was set, which is equivalent to a type 1 error of 0.05. An example of co-localisation analysis of eQTL and GWAS is shown in Figure 3-1. All significant eQTLs across all available datasets where eQTLs were found in at least one cell type or one population are displayed in Figure 3-2 and Table 3-1. In total, 49.5% (47/95) of the SLE-associated variants are found to be statistically significant eQTLs (eSNPs), which regulated the expression of 96 eGenes. By significant, I mean a FDR ≤ 0.01 and a RTC ≥ 0.95 in at least one of the data sets being tested.

When classifying the eQTLs by the distance between eSNPs and eGenes, 72% (34/47) of the significant eQTLs are cis-eQTLs and 17% (8/47) are trans-eQTLs. Of note, 11% (5/47) of the eQTLs show both *cis*- and *trans*- effects, indicating the 'mediation' effect on the *trans*-eGenes by the *cis*-mediators [160]. *Trans*-eQTLs are of special interest due to their effects on downstream genes which are not implicated by GWAS, and thereby potentially have the ability to reveal molecular pathways leading to disease [161]. These results, consistent with other studies on the genetics of gene expression [162, 163], showed a substantially higher number of genes are regulated by *cis*-eQTL rather than *trans*-eQTLs. A reason for this imbalance might be that a trans-eQTL is believed to be an indirect association and is beneath the level of detection in most studies, due to its complexity and the differences of effect size [164].

Table 3- 1. All Significant eQTLs in SLE.

eQTLs	chr	eGenes	LCL[a]	whole blood	T cells	mono.naive[b]	mono.stimul[c]	B cells	NK cells	cis/ trans
rs4649203	1	*IFNLR1*	ns	NA	4.80	ns	ns	ns	ns	cis
rs2476601	1	*PALM*	11.52	ns	ns	ns	ns	ns	ns	trans
rs1801274	1	*FCGR2A*	ns	ns	ns	ns	4.512	ns	ns	cis
rs1801274	1	*FCGR2B*	ns	ns	ns	-4.96	ns	ns	ns	cis
rs17849502	1	*WNT2*	7.52	NA	ns	NA	NA	NA	NA	trans
rs17849502	1	*LMO3*	6.98	NA	ns	NA	NA	NA	NA	trans
rs17849502	1	*IGFBP3*	7.52	NA	ns	NA	NA	NA	NA	trans
rs2297550	1	*IKBKE*	ns	ns	-5.59	5.69	5.319	ns	ns	cis
rs3024505	1	*IL10*	ns	ns	ns	ns	4.279	ns	ns	cis
rs13385731	2	*VIL1*	-6.72	NA	ns	ns	ns	ns	ns	trans
rs13385731	2	*LGALS2*	-6.96	NA	ns	ns	ns	ns	ns	trans
rs13385731	2	*RASGRP3*	ns	NA	ns	ns	ns	ns	ns	cis
rs13385731	2	*FAM98A*	ns	NA	ns	ns	-7.789	ns	ns	cis
rs2111485	2	*IFIH1*	ns	ns	ns	ns	-5.762	ns	ns	cis
rs6445972	3	*PXK*	ns	ns	6.39	ns	ns	ns	ns	cis
rs6445972	3	*ABHD6*	ns	ns	ns	-8.06	-7.434	ns	7.99	cis
rs1132200	3	*DAZL*	-5.89	ns	ns	ns	ns	ns	ns	trans
rs564799	3	*IL12A*	ns	ns	ns	ns	ns	6.02	ns	cis
rs4690229	4	*DGKQ*	4.73	44.447	4.66	28.55	26.575	24.13	ns	cis
rs4690229	4	*IDUA*	ns	13.050	ns	ns	ns	ns	ns	cis
rs4690229	4	*UVSSA*	ns	5.278	ns	ns	ns	ns	ns	cis
rs4690229	4	*CTBP1*	ns	-4.429	ns	ns	ns	ns	ns	cis
rs340630	4	*TM4SF19*	5.88	ns	ns	ns	ns	ns	ns	trans
rs7708392	5	*SLC36A1*	ns	4.883	ns	ns	ns	ns	ns	cis
rs7708392	5	*ANXA6*	ns	-6.778	ns	ns	ns	ns	ns	cis
rs36014129	6	*HAND1*	6.63	NA	ns	ns	ns	ns	ns	trans
rs36014129	6	*ITGAM*	5.97	NA	ns	ns	ns	ns	ns	trans
rs36014129	6	*TLR4*	5.76	NA	ns	ns	ns	ns	ns	trans
rs36014129	6	*GJB6*	6.30	NA	ns	ns	ns	ns	ns	trans
rs36014129	6	*TRIM36*	5.90	NA	ns	ns	ns	ns	ns	trans
rs36014129	6	*BTN3A2*	-5.73	NA	-11.76	ns	ns	ns	ns	cis
rs36014129	6	*BTN2A2*	ns	NA	4.88	ns	ns	4.14	ns	cis
rs36014129	6	*HIST1H4L*	ns	NA	-8.75	ns	ns	ns	ns	cis
rs36014129	6	*HMGN4*	ns	NA	ns	ns	4.788	ns	ns	cis
rs36014129	6	*BTN2A1*	ns	NA	ns	ns	-4.336	ns	ns	cis
rs2327832	6	*LCN8*	13.85	ns	ns	ns	ns	ns	ns	trans
rs2327832	6	*PERP*	ns	-3.619	ns	ns	ns	ns	ns	cis
rs4917014	7	*RSAD2*	ns	5.728	ns	ns	ns	ns	ns	trans
rs4917014	7	*IFI6*	ns	6.295	ns	ns	ns	ns	ns	trans
rs4917014	7	*IFI44*	ns	5.887	ns	ns	ns	ns	ns	trans
rs73135369	7	*CLEC4C*	5.67	NA	ns	NA	NA	NA	NA	trans
rs1167796	7	*PMS2P3*	ns	-5.303	ns	ns	ns	ns	ns	cis

eQTLs	chr	eGenes	LCL[a]	whole blood	T cells	mono.naive[b]	mono.stimul[c]	B cells	NK cells	cis/ trans
rs4728142	7	*TNPO3*	ns	ns	-4.20	ns	ns	ns	ns	cis
rs2070197	7	*RNF126P1*	8.28	ns	ns	ns	ns	ns	ns	trans
rs2980512	8	*NEIL2*	ns	7.704	ns	ns	ns	ns	ns	cis
rs2980512	8	*FDFT1*	ns	11.412	ns	ns	ns	ns	ns	cis
rs2980512	8	*MFHAS1*	ns	ns	ns	ns	-5.890	-6.76	ns	cis
rs2980512	8	*PRAG1*	ns	ns	ns	ns	7.418	ns	ns	cis
rs2736340	8	*FAM167A*	9.76	24.193	ns	ns	ns	23.59	ns	cis
rs2736340	8	*BLK*	-6.59	-14.310	ns	ns	ns	-13.80	ns	cis
rs2736340	8	*MTMR9*	ns	ns	ns	ns	-3.811	ns	ns	cis
rs7829816	8	*BEX2*	ns	5.316	ns	ns	ns	ns	ns	trans
rs1966115	8	*ZC2HC1A*	-8.95	ns	ns	ns	-24.242	-25.35	-4.70	cis
rs877819	10	*WDFY4*	ns	ns	ns	ns	ns	ns	ns	cis
rs4917385	10	*USMG5*	-9.40	-62.957	ns	ns	ns	ns	ns	cis
rs4917385	10	*C10ORF32*	ns	ns	ns	ns	ns	ns	ns	cis
rs4917385	10	*IFNLR1*	ns	-4.209	ns	ns	ns	ns	ns	cis
rs12802200	11	*TMEM80*	ns	ns	ns	ns	ns	ns	ns	cis
rs12802200	11	*DRD4*	ns	7.067	ns	ns	ns	ns	ns	cis
rs12802200	11	*HRAS*	ns	ns	ns	9.56	6.686	ns	ns	cis
rs494003	11	*RNASEH2C*	ns	ns	ns	ns	ns	ns	ns	cis
rs494003	11	*AP5B1*	ns	5.048	ns	ns	ns	ns	ns	cis
rs494003	11	*CTSW*	ns	-9.640	ns	ns	ns	ns	ns	cis
rs494003	11	*MUS81*	ns	ns	ns	ns	ns	-4.21	ns	cis
rs494003	11	*FIBP*	ns	ns	ns	ns	5.278	ns	ns	cis
rs4639966	11	*TMEM25*	ns	11.549	ns	ns	ns	ns	ns	cis
rs4639966	11	*CD3D*	ns	-4.731	ns	ns	ns	ns	ns	cis
rs4639966	11	*IFT46*	ns	ns	ns	ns	-3.925	ns	ns	cis
rs34330	12	*APOLD1*	ns	NA	4.96	ns	ns	ns	ns	cis
rs10774625	12	*ACAD10*	ns	NA	ns	ns	-3.786	ns	ns	cis
rs11073328	15	*FAM98B*	ns	ns	ns	ns	ns	ns	-4.90	cis
rs2289583	15	*ULK3*	ns	7.112	ns	ns	ns	ns	ns	cis
rs2289583	15	*C15ORF39*	ns	ns	ns	ns	9.012	ns	ns	cis
rs2289583	15	*FAM219B*	ns	ns	ns	ns	-4.326	ns	ns	cis
rs2289583	15	*CSK*	ns	ns	ns	5.77	ns	ns	ns	cis
rs9652601	16	*SOCS1*	ns	ns	ns	ns	-7.065	ns	ns	cis
rs16972959	16	*PRKCB*	-4.33	ns	ns	ns	ns	ns	ns	cis
rs7197475	16	*PRR14*	ns	-3.762	ns	ns	ns	ns	ns	cis
rs7197475	16	*ZNF689*	ns	ns	ns	-8.30	-3.832	-4.97	-4.21	cis
rs7197475	16	*MAPK3*	ns	ns	ns	-4.67	ns	ns	ns	cis
rs34572943	16	*DPEP3*	6.82	ns	ns	ns	ns	ns	ns	trans
rs34572943	16	*ITGAM*	ns	ns	ns	ns	-9.507	ns	ns	cis
rs34572943	16	*PYCARD*	ns	ns	ns	ns	-5.703	ns	ns	cis
rs1170426	16	*ZFP90*	-4.38	ns	ns	ns	ns	ns	ns	cis
rs2280381	16	*IRF8*	ns	ns	ns	-4.86	ns	ns	ns	cis

eQTLs	chr	eGenes	LCL[a]	whole blood	T cells	mono.naive[b]	mono.stimul[c]	B cells	NK cells	cis/ trans
rs2286672	17	*RNF167*	ns	NA	ns	ns	3.922	ns	5.27	cis
rs2941509	17	*ZPBP2*	5.18	ns	ns	NA	NA	NA	NA	cis
rs2941509	17	*IKZF3*	ns	8.120	ns	NA	NA	NA	NA	cis
rs3093030	19	*INPP4A*	5.84	NA	ns	ns	ns	ns	ns	trans
rs3093030	19	*ICAM1*	ns	NA	ns	ns	-5.023	ns	ns	cis
rs3093030	19	*ICAM4*	ns	NA	ns	-10.34	-5.502	ns	ns	cis
rs3093030	19	*ILF3-AS1*	ns	NA	ns	ns	3.627	ns	ns	cis
rs2304256	19	*TYK2*	ns	-7.009	ns	ns	ns	ns	ns	cis
rs2304256	19	*TMED1*	ns	4.313	ns	ns	ns	ns	ns	cis
rs2305772	19	*ZNF175*	-4.18	ns	ns	ns	-4.787	ns	ns	cis
rs7444	22	*UBE2L3*	7.41	26.325	ns	ns	ns	ns	ns	cis
rs61616683	22	*SYNGR1*	9.16	18.660	ns	ns	ns	7.93	15.59	cis

Number in the cell is the t-statistics from the eQTL analysis with the effect direction based on the SLE risk allele. a. t-statistics in the LCL column shows the summary results from all populations as described in Table 2-1. b. t-statistics in the 'mono.naive' column shows the summary results of naïve monocytes from Fairfax et al and Raj et al's datasets. c. t-statistics in the 'mono.stimul' column shows the summary results of all stimulated monocytes, including LPS2h, LPS24h and IFN. ns = Not significant; NA = Not Available. cis-eQTLs are located within 1MB, elsewhere are trans-eQTLs.

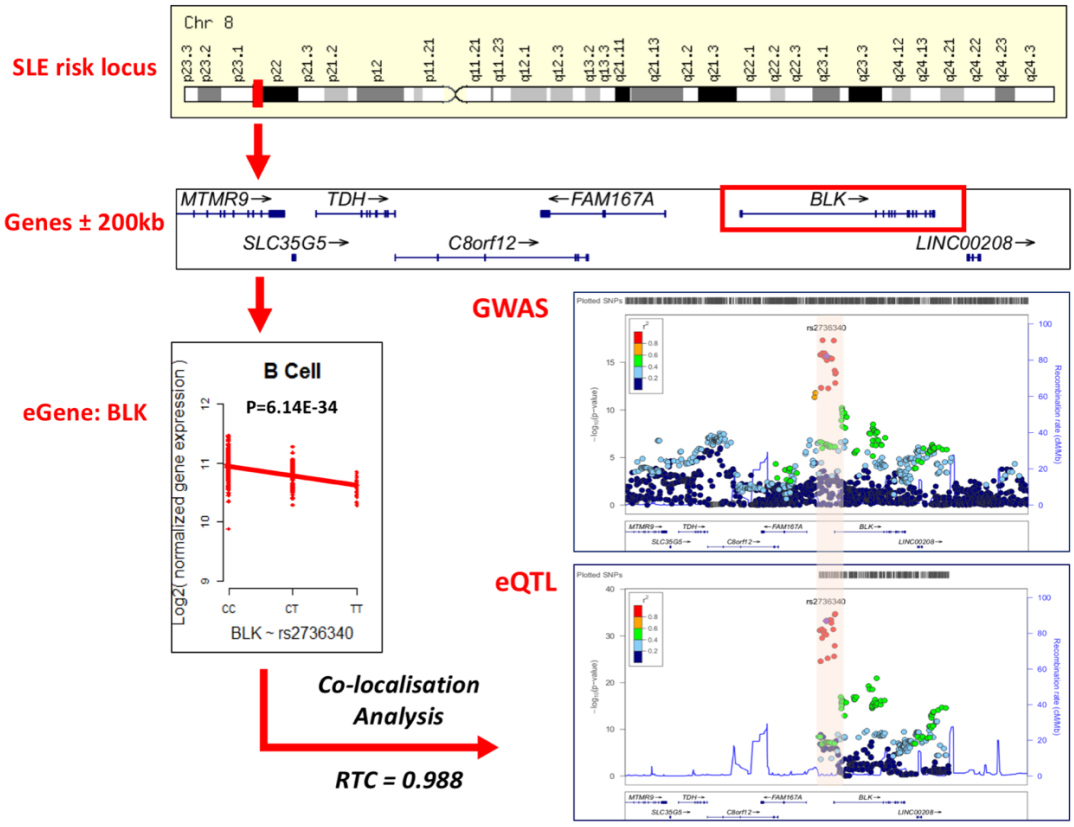


Figure 3- 1. Overview of co-localisation analysis of GWAS and eQTL.

This figure shows an example of eQTL analysis and the application of RTC for the causality inference. Firstly, I subset the genes within the cis-window (+/- 1Mb) of the disease-associated locus (rs2736340) and perform linear regression against the dosage of the SNP, coded as 0, 1, and 2, counting on the SLE risk allele (in this case, the “T” allele). Co-localisation analysis of the GWAS signal and the eQTL signal was performed by calculating the RTC score (more details in Chapter 2, Section 2.4.4). SNP-expression pairs with RTC > 0.95 were considered causal.

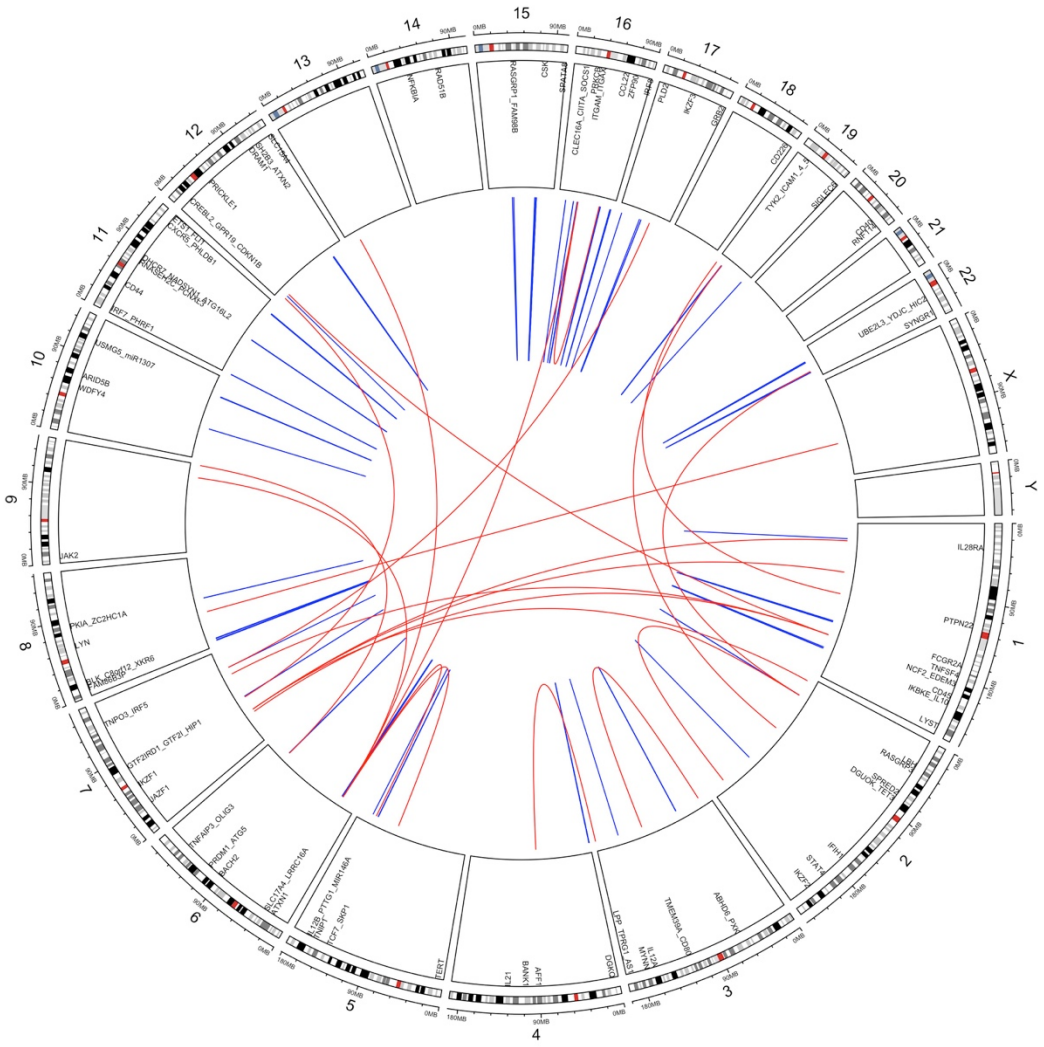


Figure 3- 2. All *cis-/trans*-eQTLs in SLE.

The outermost track of the concentric plot shows all the chromosomes clockwise. Each chromosome arm is divided into cytogenetic bands of hg19. The red block in each chromosome indicates the centromere of the chromosome. The middle layer of the concentric plot shows genes located within the SLE risk loci according to their genomic positions. The lines in the innermost of the plot indicate the significant eQTL-eGenes across the whole genome. Blue lines represent *cis*-eQTLs, while red lines represents *trans*-eQTLs . The full lists of risk loci and locus genes, as well as their eGenes are summarized in Table 3-1.

3.2 Cell-specific eQTLs

To investigate the general pattern of eQTLs across different immune cell types, I grouped the eQTLs that were identified in primary cells, including T cells, monocytes, B cells, and NK cells to gain an overview of the immune cell/cells preferences for the colocalising eQTLs. Figure 3-3 shows the SLE risk variants that have significant correlation with the expression levels of eGenes in at least one of the immune cell types. Some of the eQTLs are significantly associated with the gene expression general to all cell types, like *ZFP90* (Figure 3-4a). A significant proportion of eQTLs have their own bias for particular immune subsets, in which they regulated the target gene expression. For example, *IFNLR1* (Interferon Lambda Receptor 1, or *IL28RA*) is found to be a candidate-causal gene specific to T cells (Figure 3-4b). The IFNLR1/IL10RB dimer is a receptor for the cytokine ligands IFNL2 (Interferon Lambda 2) and IFNL3 (Interferon Lambda 3). The ligand/receptor complex stimulates the activation of the JAK/STAT signaling pathway leading to the expression of IFN-stimulated genes, which contribute to the antiviral state. Also, *IFNLR1* is reported to be associated with Psoriasis [165].

The direction of effect of eQTLs, such as *SYNGR1* (rs61616683) and *DGKQ* (rs4690229), are consistent across all cell types, with the SLE risk allele associated with increased gene expression. While eQTLs of *ZFP90* (rs1170426), *TNPO3* (rs4728142), and *ZC2HC1A* (rs1966115) are consistent with the SLE risk allele associated with reduced gene expression. For a minority of eQTLs, associations were observed across cell types with opposing direction of effect. For instance, an eSNP (rs6445972), locating in the intronic region of *PXK* in chromosome 3, is not only associated with the increased expression of *PXK* specific in T cells, but also affected the expression level of *ABHD6* (a monoacylglycerol lipase involved in the endocannabinoid system), which is significantly upregulated in NK cells, while largely down-regulated in monocytes by the SLE risk allele (Figure 3-4c). Another eSNP (rs2297550) is an eQTL for *IKBKE* (encoding IκB kinase epsilon (IKKε)), which potentially plays an important role in SLE through two underlying pathways: regulating NF-κB activation, which influences lymphocyte growth and differentiation, as well as regulating the type I interferon response, which is described in many SLE studies [166]. Cis-eQTL analyses show that the SLE risk allele is in association with the up regulation of *IKBKE* expression in monocytes, whereas it leads to a down regulation in B cell and T cell (Figure 3-4d). This highlights the molecular basis of gene regulation at different layers of complexity, cell-type specificity.

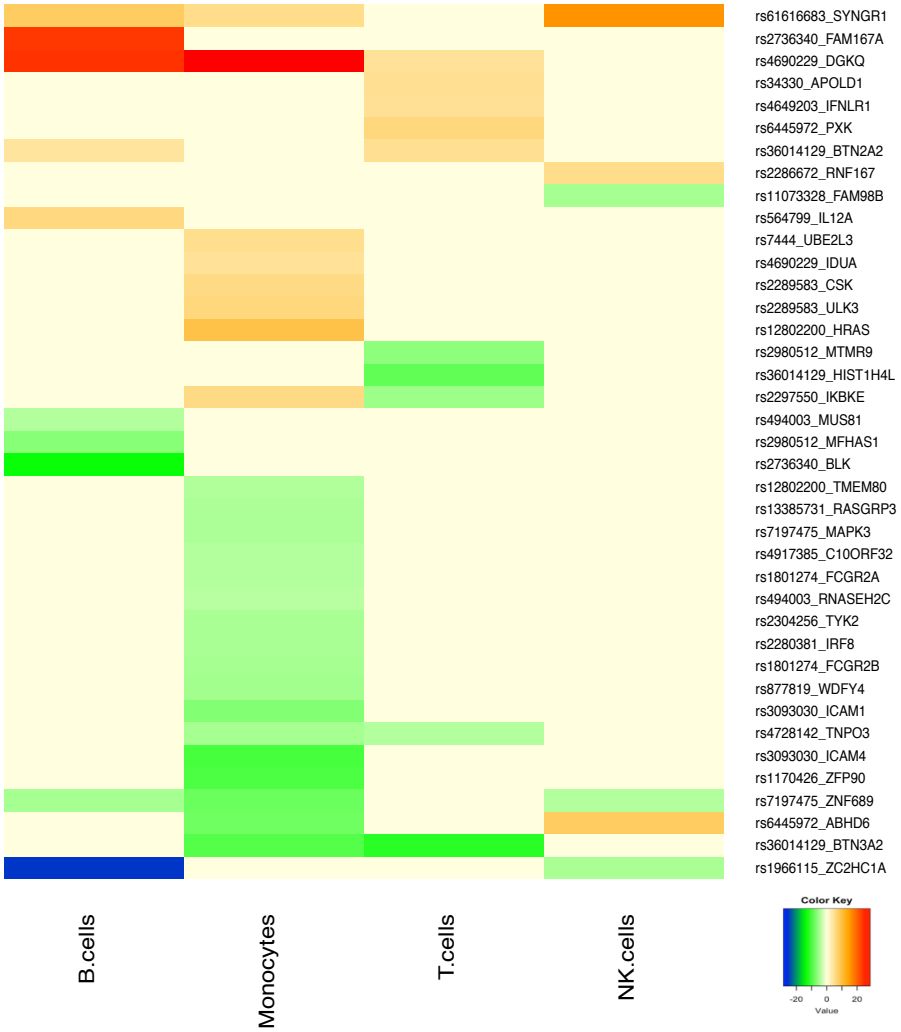


Figure 3- 3. Heat map of t-statistics for eQTLs in primary immune cells.

The heat map includes all genes with evidence of cis-regulatory action by SLE-associated SNPs in at least one of immune subsets being tested. The t-statistics was set to zero if the eQTL association is not significant (FDR >0.01 or RTC < 0.95) or the data is not available (either because the probe data failed quality control or the probe was not present on the experimental platform). For the significant eQTL associations, colours represent the t-statistics as follows: blue, t-stat < -20 (the GWAS risk allele reduces expression); green, t-stat < -10 (the GWAS risk allele reduces expression); light yellow, -10 < t-stat < 10; orange, t-stat > 10 (the GWAS risk allele increases expression); red, t-stat > 20 (the GWAS risk allele increases expression). Clustering was performed on the cell types and genes. Each row represents one gene, which is shown as "eQTL_eGene". Each column represents one cell type.

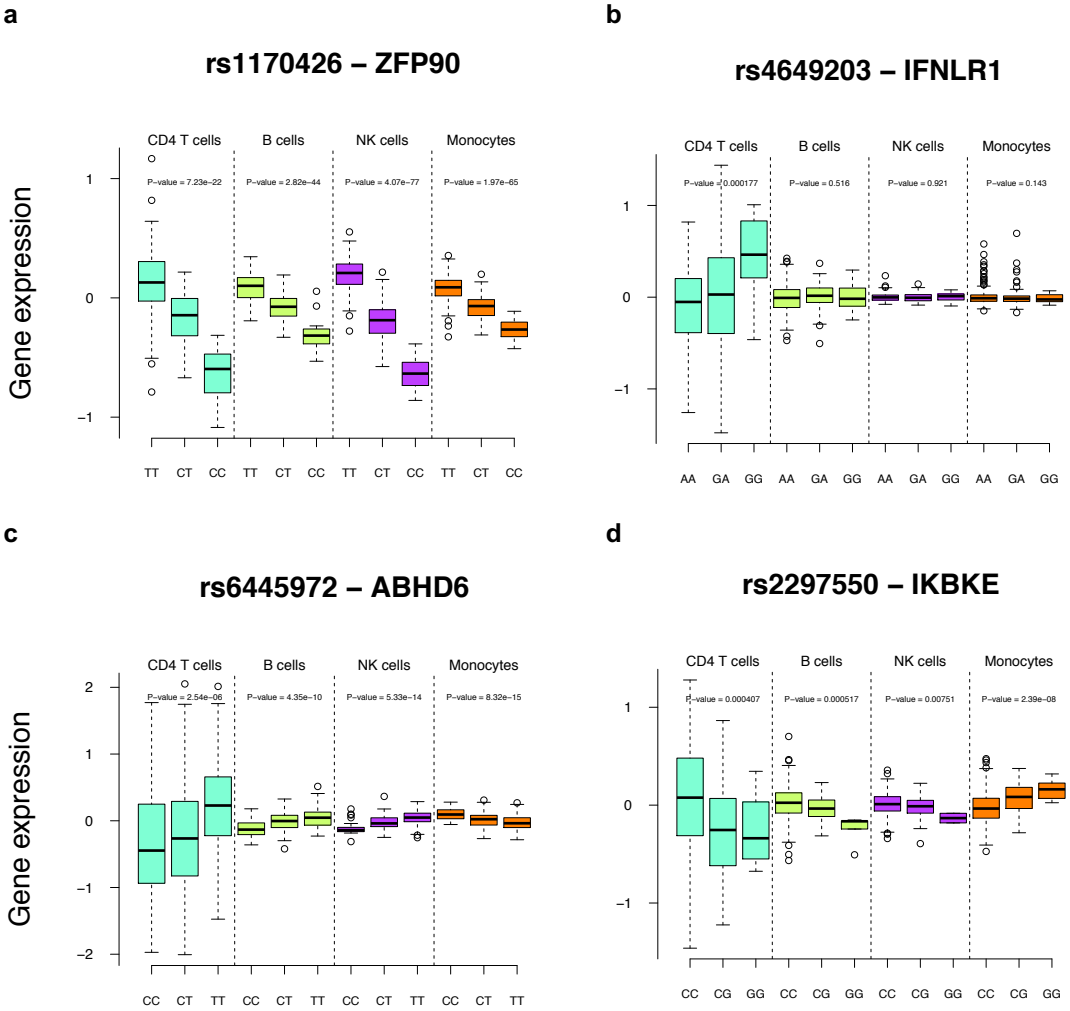


Figure 3- 4. Cell-specific eQTLs.

Boxplots of eQTLs across various immune cell types. eQTL of ZFP90 (a) is significant cross all four cell types being tested with consistent effect direction. eQTL of IFNLR1 (b) is specific to CD4+ T cells. eQTL of ABHD6 (c) has diverse effect direction across differen cell types, with an increase effect on T cells, B cells, and NK cells, while decrease in monocytes. eQTL of IKBKE (d) also shows different effect directions on immune subsets. The x-axis indicates the genotypes of the corresponding SNP and the y-axis represents the normalized gene expression. Expression data for T cells are from Raj et al [147], and gene expression data of all the other immune cells are from Fairfax et al [73].

3.3 Context-specific eQTLs in monocytes

The effect of a regulatory variant is not only dependent on the cell types as showed above in Section 4.2, but also related to the cellular and environmental context relevant to disease. In human monocytes, many regulatory variants display diverging functionality patterns of gene

regulation only after pathophysiologically relevant immune stimuli [91]. In order to better understand the regulatory role of the SLE susceptibility loci upon triggering of immune response, I used gene expression data from naïve and stimulated (IFNγ-treated 24-hour, LPS-treated 2-hour and LPS-treated 24-hour) monocytes derived from Fairfax et al [91] for eQTL analyses. Figure 3-5 shows the overall results of significant eQTLs in at least one condition of monocytes. There are some eQTLs showing strong effects on the expression of eGenes generally under all kinds of stimulations as well as the naïve state, such as *DGKQ* (rs4690229) (Figure 3-6a). There are also constitutive eQTLs present in only naïve cells but no longer present after stimulation, such as *CSK* and *IRF8*.

Nevertheless, a diverse set of immune-related genes were found to show eQTL only after induction with LPS or IFNγ, ranging from *PYCARD* (rs34572943), *IFIH1* (rs2111485), *IFT46* (rs4639966), *SIGLEC6* (rs2305772), *SOCS1* (rs9652601), to key cytokines such as *IL10* (rs3024505). For *SOCS1*, an eQTL was observed after 24-hour LPS stimulation, whereas no eQTL effect was apparent for *IL10* on induction at either 2 hours or 24 hours of LPS stimulation, but was seen an up-regulation after 24-hour IFNγ (Figure 3-6b). The expression of *PYCARD* (a key mediator in apoptosis and inflammation) was strongly correlated with rs34572943 after stimulation only after 24-hour treatment of LPS (Figure 3-6c). Notably, this eSNP (rs34572943) is located in the intronic region of *ITGAM* (encoding CD11b, part of complement receptor 3; CR3), which has been found to be associated with SLE and systemic sclerosis in many studies as a result of missense variant [45, 167, 168], suggesting the multiple effects of this locus may be involved in SLE. Of great interest, the risk allele of rs1801274 regulated the expression of multiple genes in a context specific manner, which decreased the expression of *FCGR2B* in naïve monocytes (Figure 3-6f), while increased the expression of *FCGR2A* after 2- or 24-hours LPS stimulation (Figure 3-6e). *FCGR2A* and *FCGR2B* are both encoding Fc gamma Receptor II (FcγRII), members of immunoglobulin superfamily. The process of binding FcγR with the Fc protein of Immunoglobulin G (IgG) is very important to trigger biological reactions, such as phagocytosis [169], indicating the regulatory roles of SLE risk alleles may be subjected to the disease status, such as quiescent or progressive [169, 170].

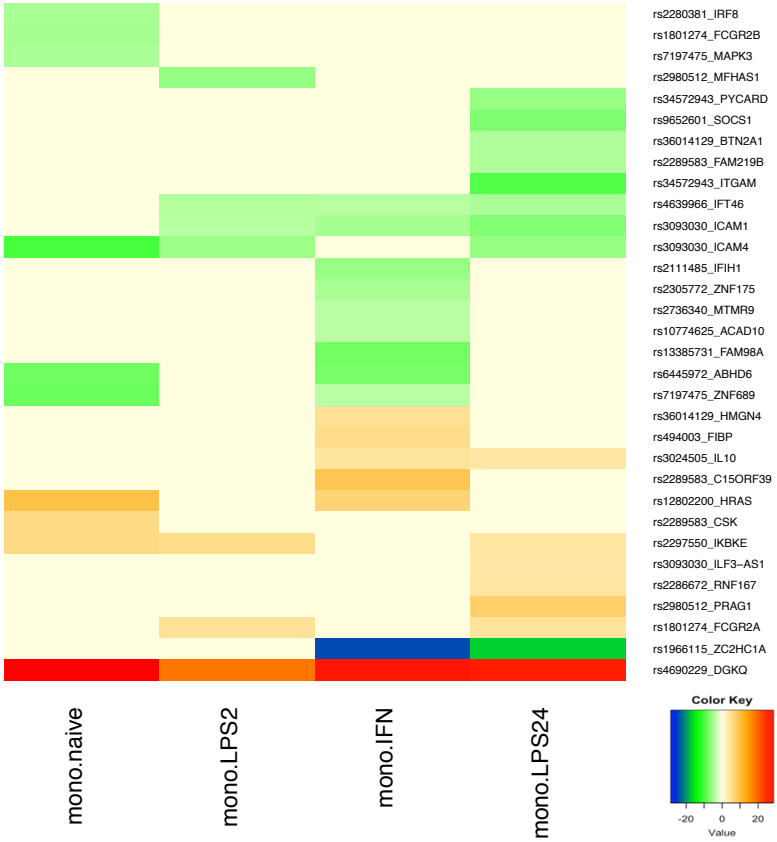


Figure 3- 5. Heat map of t-statistics for eQTLs in naïve and stimulated monocytes.

The heat map includes all genes with evidence of *cis*-regulatory action by SLE-associated SNPs in at least one status of monocytes. Colour represents a t-statistics: a positive score (red and orange) indicates that SLE risk allele is positively correlated with gene expression, and a negative score (blue and green) indicated that the SLE risk allele is negatively correlated with gene expression. The t-statistics was set to zero if the eQTL association is not significant (FDR >0.01 or RTC < 0.95) or the data is not available (either because the probe data failed quality control or the probe was not present on the experimental platform). Clustering was performed on the stimulated conditions and genes. Each row represents one gene, which is shown as "eQTL_eGene". Each column represents one cell type. Stimulated monocytes, either with lipopolysaccharide (LPS, 2 hours and 24 hours) or Interferon-γ (IFN, 24 hours). All the expression data for these analyses are derived from Fairfax et al [91].

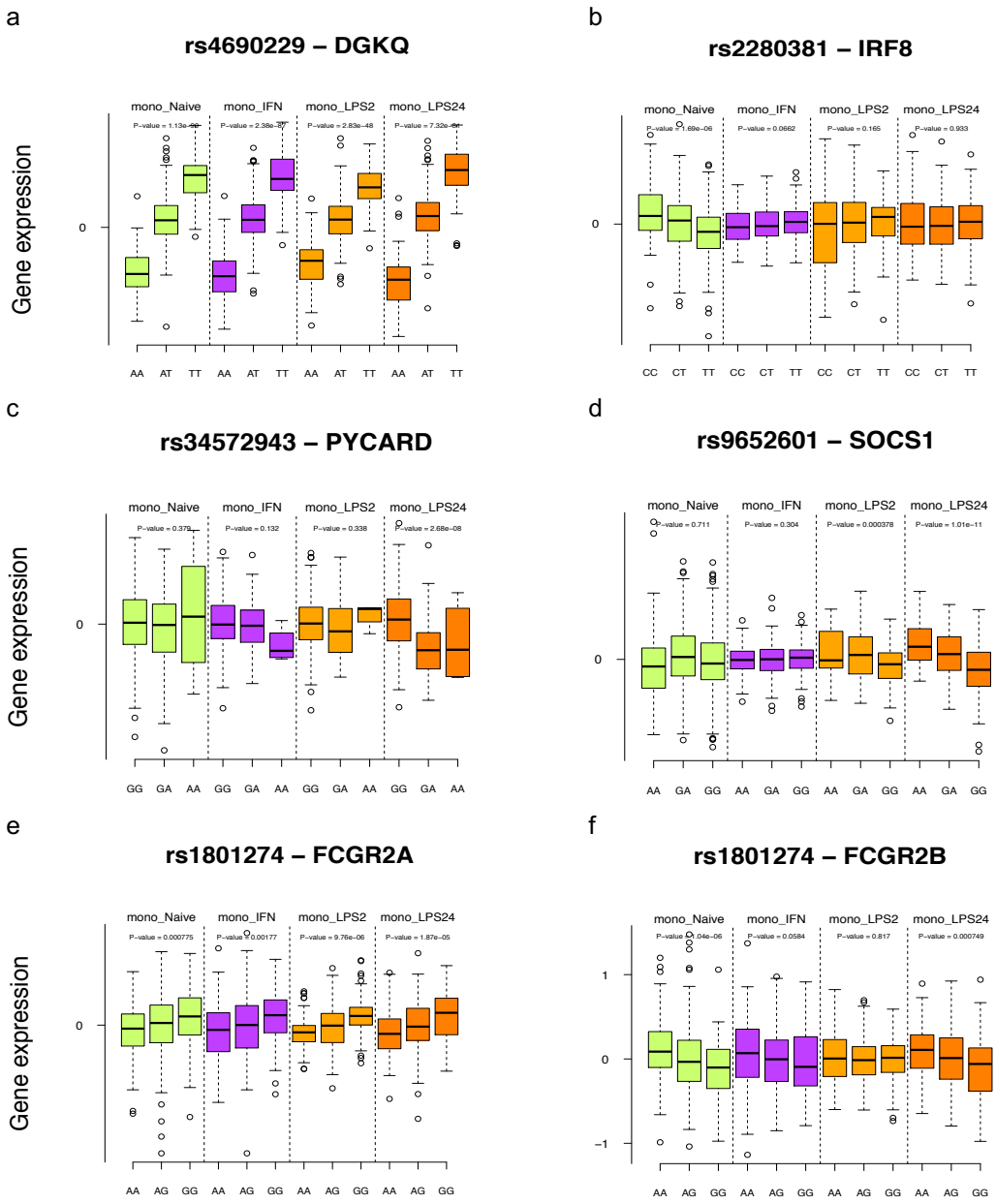


Figure 3- 6. Boxplots of context-specific eQTLs in monocytes.

Boxplots of eQTLs in monocytes with various stimulations. The x-axis indicates the genotypes of the corresponding SNP and the y-axis represents the normalized gene expression. eQTL of *DGKQ* (a) is significant cross all four conditions in monocytes. eQTL of *IRF8* (b) is only significant in naïve monocytes. Both eQTLs of *PYCARD* (c) and *SOCS1* (d) are only significant after treatment with LPS for 24 hours. rs1801274 (e and f) is an eQTL with multiple and context-specific effect.

3.4 eQTLs across populations help to infer causal eGenes

Trans-ethnic meta-analysis of GWASs across multiple populations can increase the power to detect disease associated loci if the underlying causal variants are shared between ancestry groups [54, 171, 172]. Moreover, meta-analysis of GWASs from European and Chinese

populations had identified novel loci associated with SLE [54]. Given the promising application of trans-ancestry analysis, I used data downloaded from the HapMap 3 project [149] to map the underlying causal variants in the SLE risk loci. By eQTL mapping of LCL across five populations, i.e., AFR, AMR, EUR, EAS, and SAS, I have identified significant eQTLs across different populations, see Figure 3-7. Of note, more eQTL-eGene pairs have been found in the African population, which is characterized by greater levels of genetic diversity, extensive population substructure, and less linkage disequilibrium (LD) among loci compared to non-African populations [173]. For example, rs7444 is found to be a significant eQTL for *UBE2L3* in AFR and EAS, but not in EUR. For not significant in EUR, I mean the eQTL is significant (FDR $\leq$ 0.01) but with a low RTC (RTC $\leq$ 0.95). Moreover, when analysing the rs7444-*UBE2L3* gene pair in CD4+ T cells and CD14+ monocytes across AFR, ASN and EUR in the ImmVar datasets [147], I found a similar pattern (Figure 3-8) to that observed in LCL. The RTC score is much lower in EUR, as there are many more SNPs in strong LD with the GWAS SNP within the region, than in AFR. I conclude that eQTL analyses across various populations helps to identify candidate-causal eGenes, which may be masked by the strong LD within genomic regions.

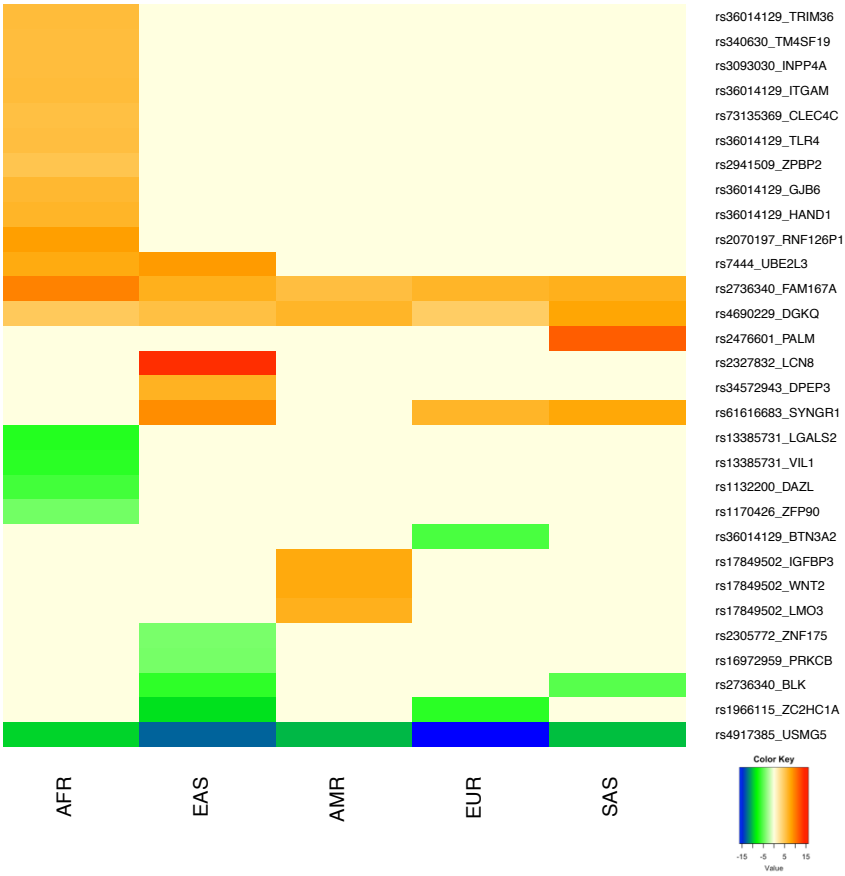


Figure 3- 7. Heat map of t-statistics for eQTLs in LCL across populations.

The heat map includes all genes with evidence of *cis-/trans-* regulatory action by SLE-associated SNPs in at least one of populations being tested. For the significant eQTL associations, colour represents a t-statistics: a positive score (red and orange) indicates that SLE risk allele is positively correlated with gene expression, and a negative score (blue and green) indicated that the SLE risk allele is negatively correlated with gene expression. The t-statistics was set to zero if the eQTL association is not significant (FDR >0.01 or RTC < 0.95) or the data is not available (either because the probe data failed quality control or the probe was not present on the experimental platform). Clustering was performed on the populations and genes. Each row represents one gene, which is shown as "eQTL_eGene". Each column represents one cell type. AFR: African; EAS: East Asian; AMR: American; EUR: European; SAS: South Asian.

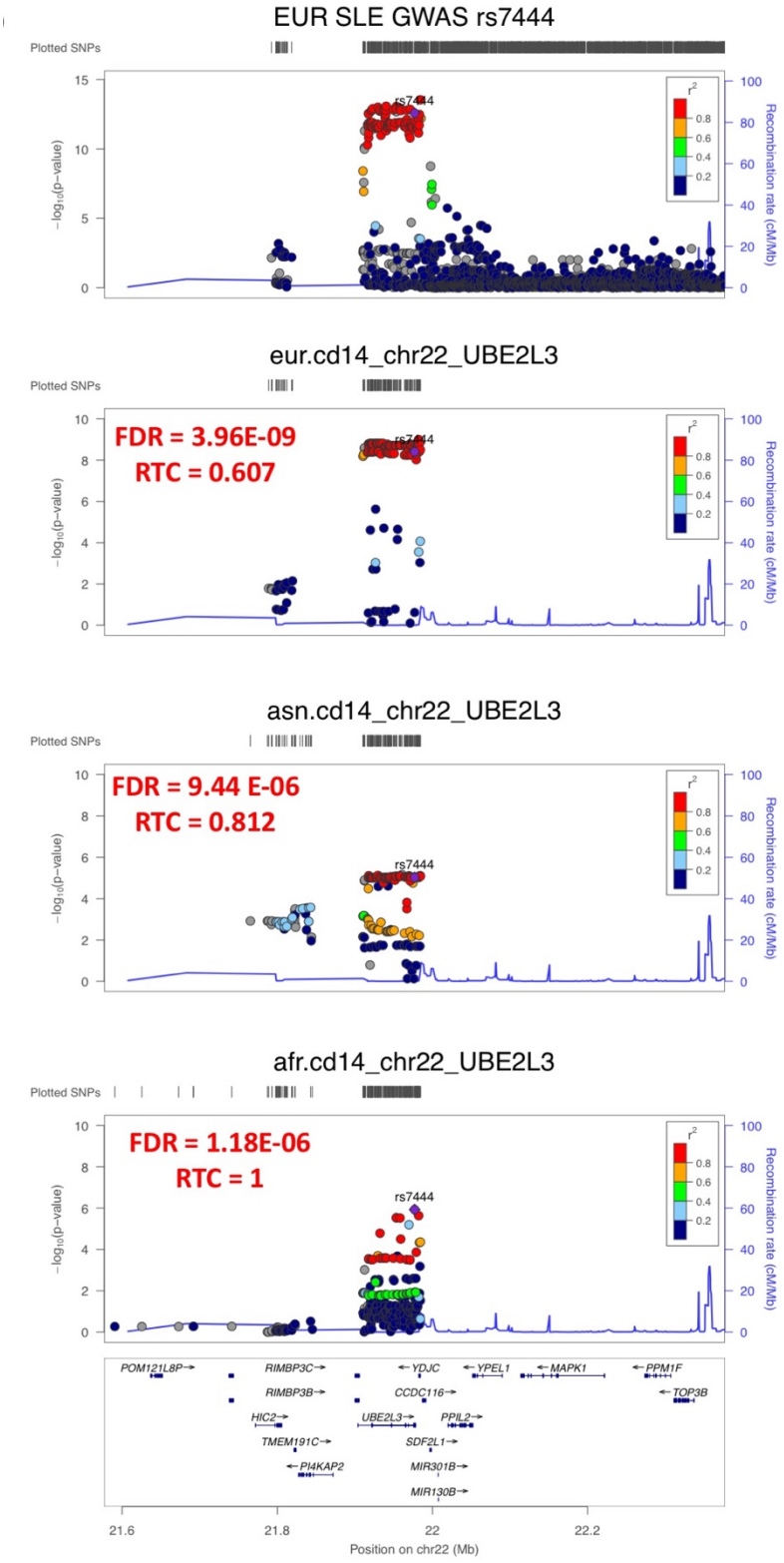


Figure 3- 8. Co-localisation analysis populations of eQTL-eGenes.

LocusZoom plots showed the co-localisation of regional P-values for SLE GWAS of the SLE main cohort [43] (top panel) and eQTLs in three populations. Data for eQTL Manhattan plot is derived from Raj et al.

3.5 eQTLs with multiple effects

Among the 47 significant eSNPs identified in this study, 21 (45%) have regulatory effects on the expression of more than one gene, resulting in a total of 71 eGenes, pointing out that the disease associated SNPs may be functionally potent in multiple genes. For example, rs494003, located in the 3'-UTR of *AP5B1* on chromosome 11, is significantly correlated with the expression of five genes, including *AP5B1, RNASEH2C, CTSW, MUS81*, and *FIBP*, with the effect directions varied across genes. Of these genes, *RNASEH2C* is of particular interest, which encodes a nucleic acid repair enzyme - Ribonuclease H2 (RNase H2), in T cells and naïve monocytes. RNase H2 degrades RNA/DNA hybrids and removes unwanted ribonucleotides from DNA. Recent research showed that RNase H2 deficiency associated with potentiation of innate immune signalling due to UV light irradiation and immune-stimulatory nucleic acid polyinosinic: polycytidylic acid [174]. The risk variant that I have identified at this locus correlates with reduced expression of *RNASEH2C* transcript. Taking together, these evidence supports a role of *RNASEH2C* in SLE pathogenesis.

Independent signals - rs7197475 and rs34572943 that are reported to be associated with SLE are located near the *ITGAM-ITGAX* region in 16p11.2. Of which, rs7197475 is found to regulate the expression of three cis-eGenes - *PRR14, ZNF689* and *MAPK3*, while rs34572943 regulated the expression of two cis-eGenes - *ITGAM* and *PYCARD*. Interestingly, *MAPK3*, encoding Mitogen-Activated Protein Kinase 3, is an essential component of the MAP kinase signal transduction pathway, whose deficiency are more potent to prime T cell response, thereby triggering the autoimmune response [175]. In my eQTL results, the SLE risk allele is correlated with declined expression of *MAPK3* in monocytes, supporting its role in balancing the immune system. Moreover, *ITGAM* and *PYCARD*, encoding key components of inflammasomes, are specific to LPS late response (Figure 3-5). These eQTL findings, taken together, provide insights for further study in the regulation network involved in this complex region (*ITGAM-ITGAX*) in terms of pathogenesis of multiple autoimmune diseases [38, 167, 176]. Other candidate-causal eQTLs with multiple functional effects include: rs1801274 is associated with the expression of both *FCGR2A* and *FCGR2B*; rs2736340 is associated with the expression of eGenes including *MTMR9, BLK, FAM167A; and* rs4690229 is associated with the expression of *DGKQ, IDUA, UVSSA,* and *CTBP1*. Of which, some eGenes could be 'passengers' and may be real in terms

of genetic effects, yet have no direct bearing on disease risk. For more multiple eQTLs, see Table 3-1.

3.6 Identifying SLE associated *trans*-eQTLs

Trans-eQTL, whose associations involving distant genes, may reveal gene networks regulated by SNPs identified by GWAS and can provide unbiased insights into pathway identity and underlying biological processes. Here I extended my eQTL analyses to include genes distant (> 1MB) from each of the target 95 SLE GWAS SNPs. With eQTL cohorts from a variety of resources, I was able to detect 13 trans-eQTLs, regulating the expression of 22 eGenes. Among these trans-eSNPs, eight of them presented trans-effects only and five of them showed both cis- and trans-effect on multiple genes. All the trans-eQTLs were identified in the whole blood cohort [148] or LCL cohort [149]. No significant trans-eQTLs were found in any of the immune subsets, which might be due to a lack of power. Notably, more than half of the trans-eQTLs were found in the African population，associating with the expression of 12 eGenes. For instances, an eSNP rs13385731, locating at the intronic region of *RASGRP3* (chr2), is significantly associated with the expression of *LAMC2 (*chr1*)* and *VIL1 (*chr2*).* While eSNP (rs73135369, in the intronic of *GTF2IRD1*) is correlated with the expression of eGene - *CLEC4C* (chr12).

An eSNP - rs36014129 (6p22.2, near the MHC), not only correlates with the expression of genes nearby - *SLC17A1, BTN3A2, BTN2A2* and, *HIST1H4L*, but also genes further away - *HAND1* (chr5), *ITGAM* (chr16), *TLR4* (chr9), *GJB6* (chr13), and *TRIM36* (chr5). Of great interest, SLE risk SNPs within the *ITGAM-ITGAX* (16p11.2) region are eSNPs for multiple eGenes, including *ITGAM*, *PYCARD*, *PRR14, ZNF689,* and *MAPK3*, as well as *DPEP3* in trans. Through pathway analysis by IPA (Ingenuity Pathway Analysis) software (Version 01-07, QIAGEN Bioinformatics), I found that the eGenes regulated by these two SLE risk loci were enriched in the pathways highly correlated with the immune functions (Table 3-2), with the top canonical pathway being the Inflammasome pathway (P = 3.13E-04). Moreover, the molecules contributing most to the enriched pathways were *TLR4, MAPK3, ITGAM,* and *PYCARD,* implying the crucial roles of these genes involved in the pathogenesis of SLE. Moreover, I have identified two trans-eQTLs in whole blood. Notably, one of them - rs4917014, locating near to *IKZF1*, is found to be a significant eQTL and affects the expression of *RSAD2* (chr2), *IFI6* (chr1), and *IFI44* (chr1). Of which, *IFI6*, encoding Interferon Alpha Inducible Protein 6, plays a critical role in the regulation of apoptosis

[177] and is involved in Interferon gamma signalling [178], suggesting its role in the pathogenesis of SLE may be crucial. All other trans-eQTLs can be found in Table 3-1 and Figure 3-9.

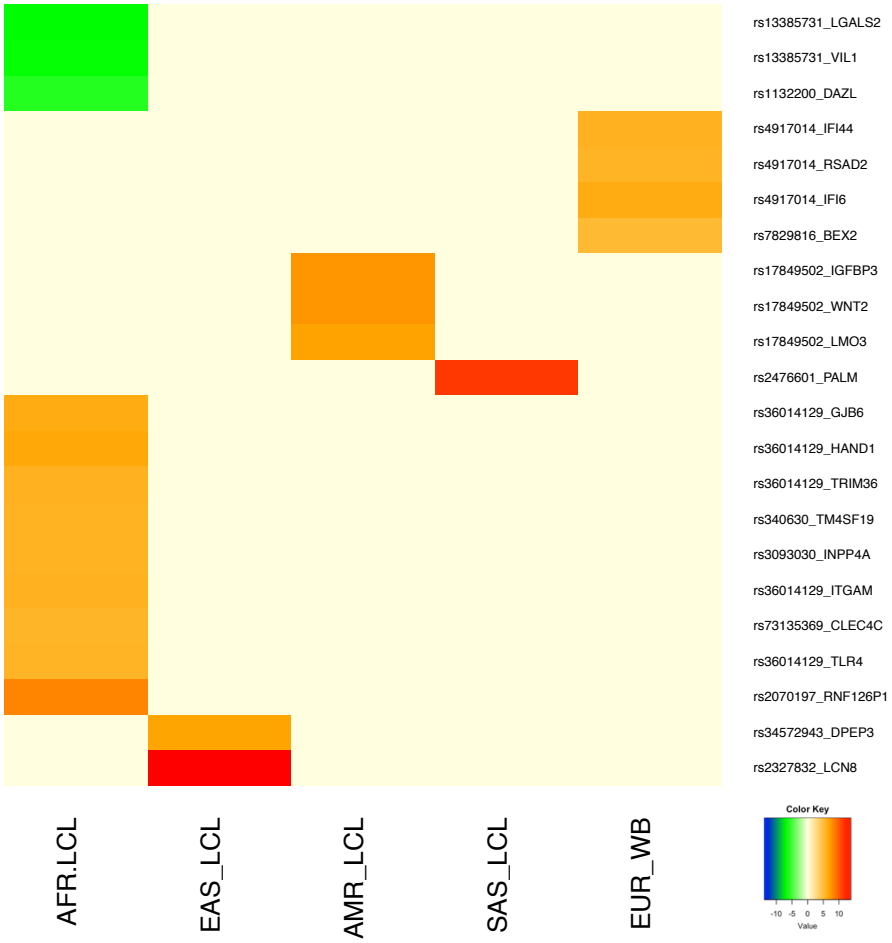


Figure 3- 9. Heat map of t-statistics for SLE associated trans-eQTLs.

The heat map includes all genes with evidence of trans-regulatory action by SLE-associated SNPs in at least one of datasets being tested. The t-statistics is set to zero if the eQTL association is not significant (FDR >0.01 or RTC < 0.95) or the data is not available (either because the probe data failed quality control or the probe was not present on the experimental platform). For the significant eQTL associations, colours represent the t-statistics as follows: blue and green, t-stat < 0 (the GWAS risk allele reduces expression); lightyellow, t-stat = 0; orange and red, t-stat > 0 (the GWAS risk allele increases expression. Clustering was performed on the cell types and genes. Each row represents one gene, which is shown as "eQTL_eGene". Each column represents one population, which is shown as “pop_cell line”.

Table 3- 2. IPA enriched pathways for eGenes regulated by risk regions in 6p22.2 and 16p11.2

Ingenuity Canonical Pathways	Ratio *	Molecules	P-value **
Inflammasome pathway	1.00E-01	TLR4; PYCARD	0.0003162
MIF-mediated Glucocorticoid Regulation	6.25E-02	TLR4; MAPK3	0.0008128
MIF Regulation of Innate Immunity	5.00E-02	TLR4; MAPK3	0.0012589
MSP-RON Signaling Pathway	3.08E-02	TLR4; ITGAM	0.0033113
TREM1 Signaling	3.08E-02	TLR4; MAPK3	0.0033113
Neuroinflammation Signaling Pathway	1.10E-02	TLR4; PYCARD; MAPK3	0.0053703
LPS-stimulated MAPK Signaling	2.35E-02	TLR4; MAPK3	0.0056234
Phagosome Formation	1.74E-02	TLR4; ITGAM	0.0100000
PI3K Signaling in B Lymphocytes	1.64E-02	TLR4; MAPK3	0.0112202
HMGB1 Signaling	1.59E-02	TLR4; MAPK3	0.0120226
IL-12 Signaling and Production in Macrophages	1.59E-02	TLR4; MAPK3	0.0120226
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	1.59E-02	TLR4; MAPK3	0.0120226
Gap Junction Signaling	1.18E-02	GJB6; MAPK3	0.0208930
Dendritic Cell Maturation	1.16E-02	TLR4; MAPK3	0.0213796
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	1.12E-02	TLR4; MAPK3	0.0229087
IL-8 Signaling	1.07E-02	ITGAM; MAPK3	0.0251189
Role of NFAT in Cardiac Hypertrophy	1.03E-02	HAND1; MAPK3	0.0269153
Integrin Signaling	1.02E-02	ITGAM; MAPK3	0.0275423
Cardiac Hypertrophy Signaling	9.62E-03	HAND1; MAPK3	0.0309030
IL-22 Signaling	4.17E-02	MAPK3	0.0316228
IL-17A Signaling in Gastric Cells	4.00E-02	MAPK3	0.0323594
Role of JAK family kinases in IL-6-type Cytokine Signaling	4.00E-02	MAPK3	0.0323594
Colorectal Cancer Metastasis Signaling	8.89E-03	TLR4; MAPK3	0.0354813
4-1BB Signaling in T Lymphocytes	3.12E-02	MAPK3	0.0416869
Complement System	3.12E-02	ITGAM	0.0416869
Oncostatin M Signaling	2.94E-02	MAPK3	0.0446684
IL-17A Signaling in Fibroblasts	2.86E-02	MAPK3	0.0457088
Thyroid Cancer Signaling	2.70E-02	MAPK3	0.0478630

Notes: A list of eGenes that are regulated by the eSNPs (SLE GWAS SNPs) in 6p22.2 and 16p11.2 regions were used for the pathway enrichment analysis. Pathways with a p-value ≤ 0.05 were listed. * Ratio indicates the number of target genes overlapped with the pathway divided by the total number of genes in the according pathway. ** P-value indicates the significance of enrichment of target genes in the pathway.

3.7 Discussion

One of the central goals of human genetics is to understand how genetic variation alters the risk of human disease and how causal genetic variants inform mechanisms underlying these diseases. Most disease-associated variants are localized in non-protein-coding regions highlighting transcriptional regulation as a common theme in the mechanism of disease predisposition. In this chapter, I integrated eQTLs and GWASs in order to offer new functional information about the aetiology of SLE. eQTL analyses, along with RTC scores, have illuminated the location and impact of regulatory variants to the expression of any given gene, thereby extending our understanding of the gene-regulatory architecture. By integrating the results of eQTL and RTC analysis, I found evidence to support the role of a single gene as candidate at a given locus. My results suggested that *SOCS1* (Suppressor of Cytokine Signalling 1) was a causal gene at the identified SLE associated SNP rs9652601 rather than *CLEC16A*, even though the risk variant resides within the latter one - a gene previously reported relating to other autoimmune diseases [179].

Currently, eQTL analyses such as these have been mostly undertaken in cell lines, predominantly EBV-transformed lymphoblastoid cell lines (LCL) [180, 181]. However, my analyses have the advantage of including gene expression data from various resources. I have shown that greater biological insight is gained from using gene expression data from immune subsets and monocytes with various treatments than using data from LCL only. By using expression data from immune subsets, I have identified cell-specific eQTLs that implied a major function of the genes in particular cell types, e.g. *BLK* in B cells and in *IFNLR1* in T cells. With eQTL mapping based on datasets of monocytes under different challenges (LPS and IFNγ), some regulatory variants display context-specific eQTL. For instances, *PYCARD* and *ITGAM* were associated with the SLE risk variants only under the stimulation of LPS, while *IFIH1* was specific to the challenge of IFNγ. Moreover, as the physiological and environmental context in patients with specific disease are of high diversity, I emphasized the importance of studying disease-specific eQTLs to more extensively resolve the functional genetic variants and the modulated genes associated with diseases.

Unlike cis-eQTL, trans-eQTL analysis performs association tests between GWAS SNPs and the expression targets in the genome-wide scale, as opposed to those located locally, which is recognized to be highly informative but challenging [161, 182]. In terms of information, trans-eQTL has the potential to point toward novel mechanistic explanations for the GWAS risk loci. On the other hand, trans-eQTL analysis is usually subjected to heavy multiple testing burdens and requires a large sample size to obtain enough power for the detection of significance. Thanks to the rich resources for the eQTL analyses for my study, I have identified several trans-eQTLs in the SLE risk regions. For instance, rs4917014, an eSNP locating near to *IKZF1*, is found to affect the expression of *RSAD2* (chr2), *IFI6* (chr1), and *IFI44* (chr1), consistent with Westra et al' study [75]. Moreover, eGenes, that are regulated by risk regions at 16p11.2 and 6p22.2, including *SLC17A3, TLR4, ITGAM, MAPK3* and *PYCARD*, may reflect novel potential regulatory networks predisposing to SLE. However, elucidating the mechanisms underpinning this observation requires further functional investigation.

The trans-ancestral genome-wide association analysis has improved the resolution to identify disease risk genes. Similarly, eQTL mapping across populations has an increasing power to detect the candidate-causal genes when incorporating with GWAS results. An overwhelming example is the identification of *UBE2L3* as a causal gene implied in SLE. *UBE2L3* has been widely studied and identified to implicate in SLE in functional studies [183]. However, analyses incorporating eQTL and RTC in a European population failed to recognize *UBE2L3* as a causal eGene (RTC = 0.6) due to a large LD block surrounding the top GWAS SNP. By eQTL mapping across populations, the 'true' causal gene was uncovered within the African population (RTC = 1), in which, a narrow peak of eQTL signal was revealed (Figure 3-8).

Although I have identified some intriguing eQTLs which are cell-specific or context-specific in this study, I may neglect other mild to modest eQTLs due to the relatively small sample size of the available datasets. Take the Ikaros family of genes as examples, I failed to identify *IKZF2* or *IKZF3* as eGenes with the currently available datasets. Nevertheless, the expression of *IKZF1/IKZF2/IKZF3* were all found to be significantly correlated with their corresponding SLE GWAS signals in the Framingham Heart Study cohort, a whole blood dataset [85] with a sample size of over 5,000. In addition, with the development of next generation sequencing (NGS) technologies, the resolution on measuring gene expression will be largely improved and the power to detect eQTLs will be undoubtedly extended. Studies by my colleague - Odhams et al [184, 185] have shown that by profiling RNA-Seq data at multiple resolutions, more eQTLs had

been observed particularly at the exon-level. Though expression data generated from MicroArray are a good surrogate for measuring transcript at probe level, using RNA-seq to quantify transcript level at multiple resolutions, including gene-, exon-, splice-junction, and isoform, will very likely advance our understanding of the implications of the genetic control of gene expression. Therefore, I emphasize that large RNA-seq datasets for eQTL studies should be generated across trait-associated cell types and under disease conditions.

In summary, I have identified cell-specific and context-specific eQTLs by analysing data generating from a variety of immune subsets and monocytes triggered with different stimulations. By mapping eQTL across populations in LCL, I demonstrated the effectiveness of trans-ethnic eQTL mapping to identify potential causal genes. Furthermore, I have identified trans-eQTLs, providing evidence for further functional studies to elucidate the underlying regulatory networks implicated in the pathogenesis of SLE.

Chapter 4. The functional analysis of SLE genetics

Chapter Summary

SLE is a complex autoimmune disease with a variety of organs and tissues involved, including the central nervous system, cutaneous complications, pleural effusion, pericarditis effusion, renal and gastrointestinal manifestations. The range of symptoms of SLE can make the diagnosis complicated. The likelihood of developing SLE has a genetic component and this genetic risk is inherited in a polygenic manner, in most cases. The advent and application of genome-wide association studies (GWASs) over the last decade has dramatically advanced understanding of the genetics of SLE. The explained heritability by genome-wide associated variants was reported to be 15% in the largest European population in 2015 by our group [43]. This study increased the number of SLE associated loci to 48. In 2016, an Asian [ASN] GWAS [50] and a meta-analysis between European [EUR] and ASN GWAS [54] were performed, and more novel loci were discovered. Currently there are over 80 genomic loci robustly associated with SLE [64]. Genome-wide typed variants in both Chinese and European GWAS explain 28% of the heritability of SLE (chip heritability), with the target being between 42% – 66% [11]. Despite considerable success in identifying many SLE associated loci, the functional effects of the risk alleles for the majority of these loci remain to be defined. Up to 90% of the risk variants are located outside the protein coding regions. How these variants predispose to the disease and how these variants influence the phenotypes of patients remain unknown. In this section, I explore the functional context of SLE associated variants by integrative analysis of epigenomic data from a variety of tissues and cell types in the human body.

Firstly, I reanalysed the SLE associated signals in Europeans using re-imputed data. Then I applied a non-parametric method, GARFIELD [144], along with regulatory epigenomes obtained from ENCODE and Roadmap to test whether the SLE associated loci are enriched in DNase hypersensitivity sites (DHS) and, if they are enriched, in what cell types/tissues. I also use another algorithm, GoShifter [146], to find out what loci are contributing to the enriched cell type/tissue of DHS. Moreover, by comparing if the SLE risk allele was associated with both gene expression and DHS hotspot, I found that 66% of the SLE eQTLs were shown to overlap with DHS hotspots, denoting extensive local coordination of genetic influences on gene expression and epigenetics.

To my knowledge, this is the first comprehensive study of the regulatory map of SLE risk loci across human cells and tissues. I conclude that the SLE risk loci are primarily enriched in regulatory regions within a range of different immune cells and a large proportion of them were also shown to regulate to the expression of genes nearby.

4.1 SLE GWAS findings

My colleagues in my research group have re-imputed the genotyping data derived from the Illumina 1M Chip used to genotype 4,036 European SLE patients in the published GWAS [1]. This re-imputed data, using the latest reference genome (Phase 3) released by the 1000 Genome Project [74], generating more than 80 million variants, has not yet been published. SLE GWAS based on the high density imputed genotypes were performed using SNPTEST [141] by colleague David Morris. SNPs with imputation INFO scores of <0.7 and MAF < 0.001 were removed from further analysis. After quality control (QC), there are 21,431,070 SNPs left for further analysis. In total, there are 19,604 SNPs that reached genome-wide significance (p-value $\leqslant 5 \times 10^{-8}$) in the association test, which mapped to 63 loci, with the MHC region defined as one extended region of 25Mb ~ 35Mb in chromosome 6. For regions outside the MHC, the distance between SNPs inside the same locus is less than 1Mb. Among the 62 non-MHC loci, 34 loci have more than one SNP that passed the genome-wide significance level due to LD, while the remaining 28 loci have only one genome-wide significant SNP in their corresponding loci. Table 4-1 shows all the loci and their lead GWAS SNPs. I name these loci according to their positions (i.e. name of gene nearby). Summary GWAS results genome-wide are shown in the Manhattan plot (Figure 4-1). Loci that have previously been shown to be significantly associated with SLE can be seen in the GWAS paper by colleagues Morris and Sheng et al [54].

Table 4- 1. Summary of SLE risk SNPs for each chromosome.

Chromosome ID	# QCed SNPs *	# Significant SNPs **
1	1665797	288
2	1827613	229
3	1537201	10
4	1526428	122
5	1375499	75
6	1425495	17783
7	1233388	265
8	1204366	59
9	931227	0
10	1068362	23
11	1062426	196
12	1011657	2
13	784755	0
14	686487	1
15	606576	0
16	642710	293
17	554393	2
18	594170	0
19	435408	20
20	458937	17
21	280729	0
22	269508	112
X	247938	107
Total	**21431070**	**19604**

*: # QCed SNPs denoted the number of SNPs that passed QC criteria - imputation info score ≥ 0.70 and minor allele frequency (MAF) ≥ 0.001. **: # Significant SNPs denoted the number of SLE risk SNPs that had a genome-wide significant P-value $\leq 5 \times 10^{-8}$.

Table 4- 2. SLE GWAS summary loci of the SLE main cohort.

LocusGenes	topSNP	chr	pos	alleleA	alleleB	Func.refGene	MAF	INFO score	OR	P value	Notes
KAZN	rs4661543	1	15229101	T	G	intronic	0.12392	1	1.3282549	9.399E-11	balloon*
NFIA-MGC34796	rs12081621	1	61943156	C	T	intergenic	0.0034561	1	6.5976927	8.716E-13	balloon
PTPN22-RSBN1	rs2476601	1	114377568	A	G	exonic	0.101273	1	0.7182382	8.382E-13	
CFAP126-FCGR2A-HSPA6-FCGR3A	rs6671847	1	161478810	G	A	intronic	0.487569	0.994824	1.2219134	3.594E-12	
TNFSF4-LOC100506023	rs10912578	1	173251856	A	G	ncRNA_intronic	0.329585	0.984176	0.784179	2.239E-15	
SHCBP1L-LAMC1-NMNAT2-SMG7-AS1-NCF2-ARPC5	rs17849501	1	183542323	C	T	exonic	0.0773079	1	2.2427034	1.814E-59	
ITGB1BP1	rs4074976	2	9546695	G	T	UTR3	0.0080946	1	2.8670109	2.714E-11	balloon
SPRED2-MIR4778	rs268124	2	65654364	C	T	intronic	0.253474	0.992707	1.2116596	7.64E-09	
IL37-IL36G-IL1F10	rs13019891	2	113829869	G	T	intronic	0.399909	1	0.5688474	1.647E-83	
LOC101927709-EN1	rs512681	2	119450163	C	A	intergenic	0.139245	1	0.7424001	8.749E-13	balloon
NEMP2-NAB1-GLS-STAT1-STAT4	rs141442738	2	191955189	TATA	T	intronic	0.256394	0.99473	1.756115	1.157E-66	
IKZF2	rs10048743	2	213890232	G	T	intronic	0.138199	1	0.7932888	2.038E-08	
KCNE4-SCG2	rs10200680	2	223961877	C	T	intergenic	0.132924	1	0.7794981	4.963E-09	balloon
ALPPL2-ALPI	rs2573219	2	233288667	A	C	intergenic	0.115098	1	1.8083141	1.133E-42	balloon
SATB1-AS1-KCNH8	rs11928304	3	18998569	A	G	intergenic	0.0056389	1	32.939416	1.95E-48	balloon
PXK-PDHB-KCTD6	rs11719330	3	58462163	G	A	intergenic	0.135823	0.87707	0.7615321	2.287E-09	
H1FX-AS1-RPL32P3	rs9852014	3	129084581	A	G	intergenic	0.0940882	1	1.8576625	2.257E-36	balloon
PLSCR5-LINC02010	rs1464446	3	146601295	G	T	intergenic	0.160709	1	0.7216572	2.792E-16	balloon
IQCJ-SCHIP1-IL12A-AS1-IFT80-TRIM59-KPNA4	rs80014155	3	159625393	C	T	intergenic	0.0105516	0.947037	2.7678753	7.383E-13	
MIR1269A-LOC101927237	rs17087866	4	67876041	A	G	intergenic	0.0043656	1	9.4240035	2.591E-22	balloon
PPP3CA-BANK1	rs6532924	4	102239170	T	C	intronic	0.0048204	1	5.7565596	8.579E-17	
FBXL7	rs13170409	5	15764650	A	C	intronic	0.0028649	1	5.4523134	7.854E-10	balloon
CDH18	rs141208658	5	20469875	T	C	intronic	0.0025154	0.778681	6.8943346	2.794E-08	
ANKRD31	rs183086601	5	74366599	T	C	intronic	0.0028866	0.83443	9.276E-06	5.784E-11	balloon
VDAC1-TCF7	rs6899047	5	1.33E+08	G	C	intergenic	0.0535808	0.991903	1.4792364	3.177E-10	
PPARGC1B	rs1078324	5	1.49E+08	C	A	intronic	0.0407913	1	0.4859833	7.105E-20	balloon
TNIP1	rs6889239	5	1.5E+08	T	C	intronic	0.274556	0.999226	1.3250007	1.205E-18	
PTTG1-MIR3142HG	rs2431697	5	1.6E+08	T	C	intergenic	0.422738	1	0.8011076	2.601E-14	
E2F3	rs11962557	6	20407282	G	A	intronic	0.002683	1	13.892373	3.542E-18	
MHC	7 independent SNPs	6									
B3GAT2-OGFRL1	rs9969061	6	71862393	A	G	intergenic	0.0020919	1	13.297951	1.118E-13	
PRDM1-ATG5	rs7768653	6	1.07E+08	C	T	intergenic	0.400953	0.994895	0.8127246	1.575E-12	
LOC102724152-MEAT6	rs16895550	6	1.64E+08	T	C	intergenic	0.0039563	1	0.1116697	5.99E-11	
TNFAIP3-PERP-LOC100507406-LOC100130476	rs148314165	6	1.38E+08	GT	G	intergenic	0.0381637	0.985315	1.8423649	8.894E-17	
DGKB-AGMO	rs17168663	7	15139469	T	G	intergenic	0.0025011	1	7.1856789	7.767E-12	balloon
KCP-IRF5-TNPO3-TPI1P2	rs12531711	7	1.29E+08	A	G	intronic	0.13271	0.998103	1.8017673	7.065E-46	
PLXNA4-FLJ40288	7:132339127	7	1.32E+08	T	C	ncRNA_intronic	0.0006517	0.977564	66078.585	2.764E-08	
AKR1B15-BPGM	rs10264693	7	1.34E+08	A	G	intergenic	0.0046385	1	43.298587	2.311E-42	balloon
PPP1R3B-LOC101929128	rs106380	8	9035078	G	C	intergenic	0.433186	0.988578	1.1748201	3.573E-08	balloon
XKR6-MTMR9-FAM167A-BLK	rs2736332	8	11339965	G	C	intergenic	0.293333	0.979054	1.3212708	1.351E-18	
LOC100507464-SNTG1	rs7386188	8	50609940	T	G	intergenic	0.0297863	1	1.650587	1.105E-08	balloon
SOX17-RP1	rs7823055	8	55511676	G	T	intergenic	0.449022	1	0.7080554	1.643E-34	balloon
LINC00708-LOC105376398	rs71287391	10	8446489	AT	A	intergenic	0.300557	0.774732	0.8185547	2.349E-08	
MAP3K8-LYZL2	rs1034009	10	30757105	C	A	intergenic	0.0038199	1	5.8853723	7.129E-13	balloon
WDFY4	rs7097397	10	50025396	G	A	exonic	0.372105	0.978192	0.8236859	1.061E-10	
MIR4682-PLPP4	rs12255478	10	1.22E+08	T	C	intergenic	0.004443	0.930832	0.0552299	8.952E-12	balloon
RNH1-HRAS-LRRC56-MIR210HG-IRF7-PHRF1-CDHR5-SCT-DRD4-DEAF1	rs4963140	11	588654	C	T	intronic	0.387644	0.995366	1.2210022	1.326E-11	
PDHX-LOC100507144	rs353608	11	35101738	A	G	intergenic	0.455255	0.996157	1.2145104	2.042E-11	
PIK3C2G	rs12309414	12	18665717	A	G	intronic	0.0050478	1	29.327628	1.804E-42	balloon
HECTD4	rs7953257	12	1.13E+08	A	T	intronic	0.432067	0.933497	0.8486359	4.072E-08	balloon
PELI2-LOC101927690	rs17091347	14	56820049	T	C	intergenic	0.0043656	1	4.649176	8.662E-13	balloon
PRSS36-FUS-PYCARD-AS1-TRIM72-ITGAM-ITGAX	rs8056944	16	31283744	C	G	intronic	0.159642	0.998965	1.7424776	2.871E-47	
PRSS54-GINS3	rs140819743	16	58342077	AT	A	intergenic	0.303554	0.93452	0.8354849	1.883E-08	balloon
IRF8-LINC01082	rs8052690	16	85969738	A	G	intergenic	0.207513	0.988756	0.7373202	5.987E-17	
MRM1-LOC102723471	rs8078864	17	35056109	T	C	intergenic	0.0069122	1	7.4873131	8.929E-29	balloon
GPATCH8	rs12948819	17	42521824	A	G	intronic	0.0033197	1	17.920226	2.982E-24	balloon
ICAM5-ZGLP1-FDX1L-ICAM3-TYK2-CDC37-PDE4A	rs11085727	19	10466123	C	T	intronic	0.272683	0.980055	0.7895406	6.911E-13	
LINC01751-LINC01706	rs56412650	20	7333291	C	T	ncRNA_intronic	0.0025214	0.976837	5.8133675	5.422E-09	
CHMP4B-RALY-AS1-EIF2S2-ASIP-AHCY-ITCH	20:32515626	20	32515626	G	A	intergenic	0.0011236	0.80031	45.627016	2.452E-10	
UBE2L3-YDJC	rs181361	22	21929566	A	T	intronic	0.212994	0.992962	1.2935754	9.716E-14	
MIR3201-FAM19A5	rs7285053	22	48872193	T	C	intergenic	0.0043656	1	14.999397	4.038E-29	balloon
NR0B1-CXorf21	rs2532871	X	30573733	G	C	intergenic	0.0146835	0.195635	8.436E+11	1.161E-08	balloon
ARHGAP4-NAA10-RENBP-HCFC1-TMEM187-IRAK1-MECP2-OPN1LW	rs13397	X	1.53E+08	G	A	exonic	0.165917	0.925831	1.6750282	1.774E-13	

*balloon, there is only one genome-wide significant SNP in the region.

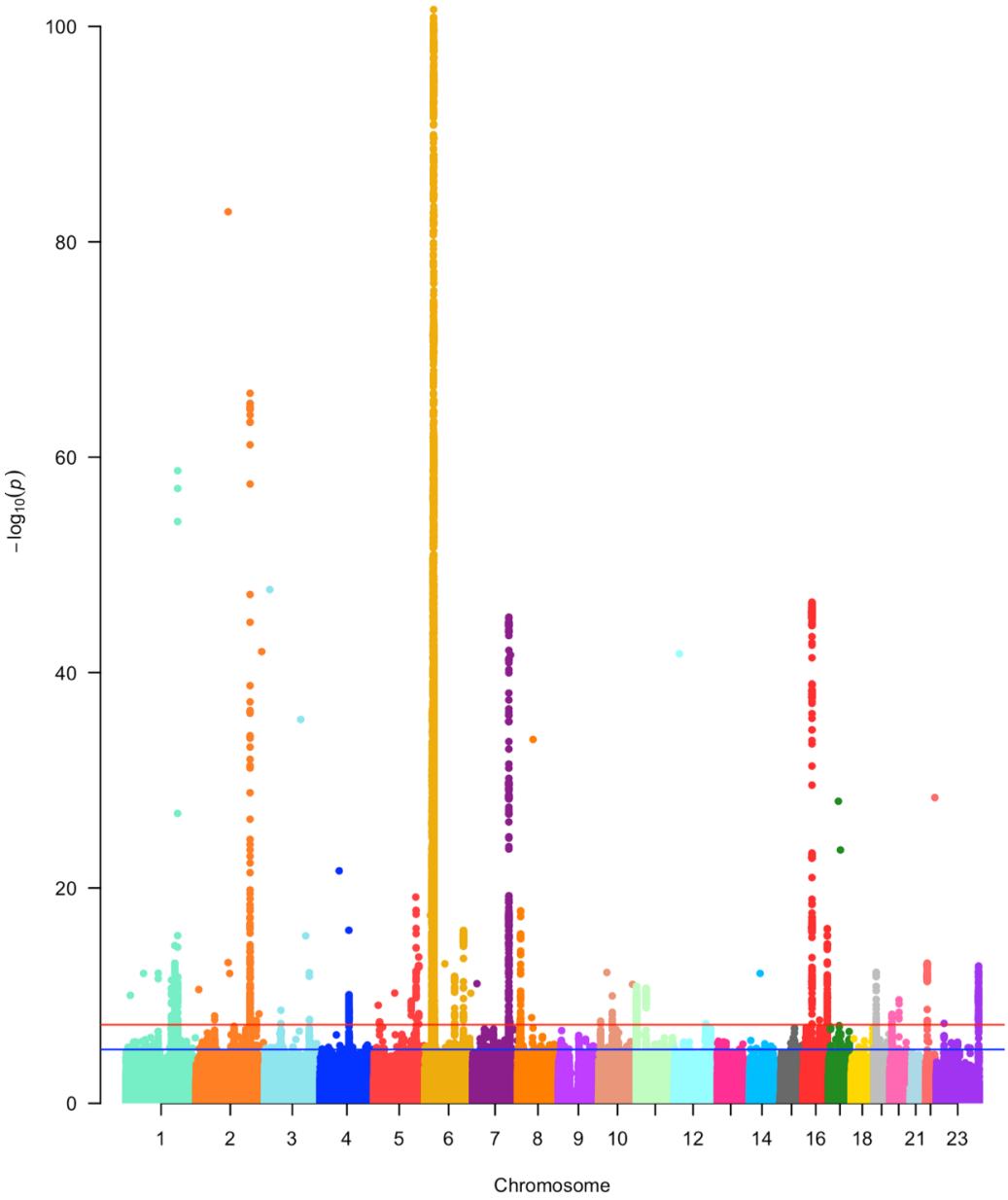


Figure 4- 1. Manhattan plot of SLE GWAS results in SLE main European cohort.
The GWAS identifies 63 loci ($P \leqslant 5 \times 10^{-8}$) that are associated with SLE. The $-\log_{10}$ (P values) derived from the Wald statistic (logistic regression model) were plotted against chromosomal position. Each point represents a SNP. The red line corresponds to the genome-wide significance threshold ($P=5 \times 10^{-8}$). The blue line corresponds to $P=1 \times 10^{-5}$.

4.2 Enrichment pattern of SLE risk loci

The majority of the disease risk variants identified from the European SLE GWAS (and the other SLE GWAS) are located outside the protein coding regions of the genome. Only four out of the 62 loci have a top SNP that was located in an exonic region, while the remaining are intronic,

intergenic, or UTR, where the functional effects of the risk alleles are in general, poorly defined. To evaluate the functional properties of these variants, I estimated the extent to which these SLE associated variants were non-randomly distributed across various coding, non-coding regulatory and cell-type-specific elements across the genome. 424 genome-wide data sets derived from the most widely studied annotation type, DNaseI hypersensitivity sites (DHS) hotspots, from the ENCODE [94] and Roadmap projects [96] were retrieved for analyses (See Chapter 2. Methods). A nonparametric approach, GARFIELD [143], was applied to derive fold of enrichment statistics for SLE risk variants within each cell type/tissue, where SNPs were selected on the basis of their strength of association with SLE (GWAS P value). Following the enrichment analysis, I found that genome-wide significant SLE associations ($P = 1 \times 10^{-8}$) are significantly (with an empirical $P < 1 \times 10^{-4}$ and Bonferroni adjusted $P < 0.05$) enriched in the DHS hotspots of immune cells involving in the adaptive immune response, including CD3 T cells (fold enrichment = 3.95), CD4 T cells (fold enrichment = 2.81), CD8 T cells (fold enrichment = 3.34), CD19 B cells (fold enrichment = 3.81), and CD20 B cells (fold enrichment = 3.46), as well as cells that play a role in the innate immune response, comprising CD56 NK cells (fold enrichment = 3.54) and CD14 monocytes (fold enrichment = 3.05). Other cells and tissues that are significantly enriched for SLE risk loci include CD34 hematopoietic stem cells, lymphoblastoid cell lines (LCL) (fold enrichment = 3.57), fetal large intestine (fold enrichment = 2.59), fetal small intestine (fold enrichment = 2.40), fetal spleen (fold enrichment = 1.89), and breast (fold enrichment = 1.93). See Figure 4-2 and Table 4-3.

Table 4- 3. Summary statistics of SLE risk loci enrichment analysis by GARFIELD.

Cell type/tissue	N	N_a	N^t	N_a^t	FE	Empirical *P* value	Adjusted *P* value
CD19+ B cells	295822	24801	288	92	3.81028	<0.0001	<0.05
CD20+ B cells	295822	29425	288	99	3.45586	<0.0001	<0.05
CD3+ T cels	295822	21846	288	84	3.94953	<0.0001	<0.05
CD4+ T cells	295822	22626	288	62	2.81463	<0.0001	<0.05
CD8+ T cells	295822	18736	288	61	3.34419	<0.0001	<0.05
Th2	295822	23486	288	55	2.40542	<0.0001	<0.05
CD14+ Monocytes	295822	23602	288	70	3.0464	<0.0001	<0.05
CD56+ NK cells	295822	19133	288	66	3.54323	<0.0001	<0.05
CD34+ stem cells	295822	26574	288	64	2.47378	<0.0001	<0.05
LCL	295822	23283	288	81	3.57342	<0.0001	<0.05
Large Fetal Intestine	295822	22243	288	56	2.58602	<0.0001	<0.05
Small Fetal Intestine	295822	26923	288	63	2.40356	<0.0001	<0.05
Fetal Spllen	295822	37979	288	70	1.89318	<0.0001	<0.05
Breast (vHMEC)	295822	31864	288	60	1.93414	<0.0001	<0.05

Notes: *N* is the total number of pruned SNPs; N_a is the total number of SNPs that fall in DHS regions (functional annotation); N^t is the total number of variants with SLE GWAS P-value less than threshold t (in this table is 1×10^{-8}); N_a^t denotes the number of SNPs with P value less than t and fall within DHS. FE (fold of enrichment) is the proportion of the SNPs that fall in DHS as well as have a P-value less than threshold t, divided by the fraction of total number of SNPs within DHS, i.e., FE = $(N_a^t/N^t)/(N_a/N)$. Empirical *P* value is the proportion of times a random selection of SNPs would have a higher FE than observed, which is calculated by 10,000 times of permutations. Adjusted *P* value is the empirical *P* value accounting for multiple-testing burden by Bonferroni adjustment. For more details, refer to Chapter 2, 2.3.1 GARFIELD.

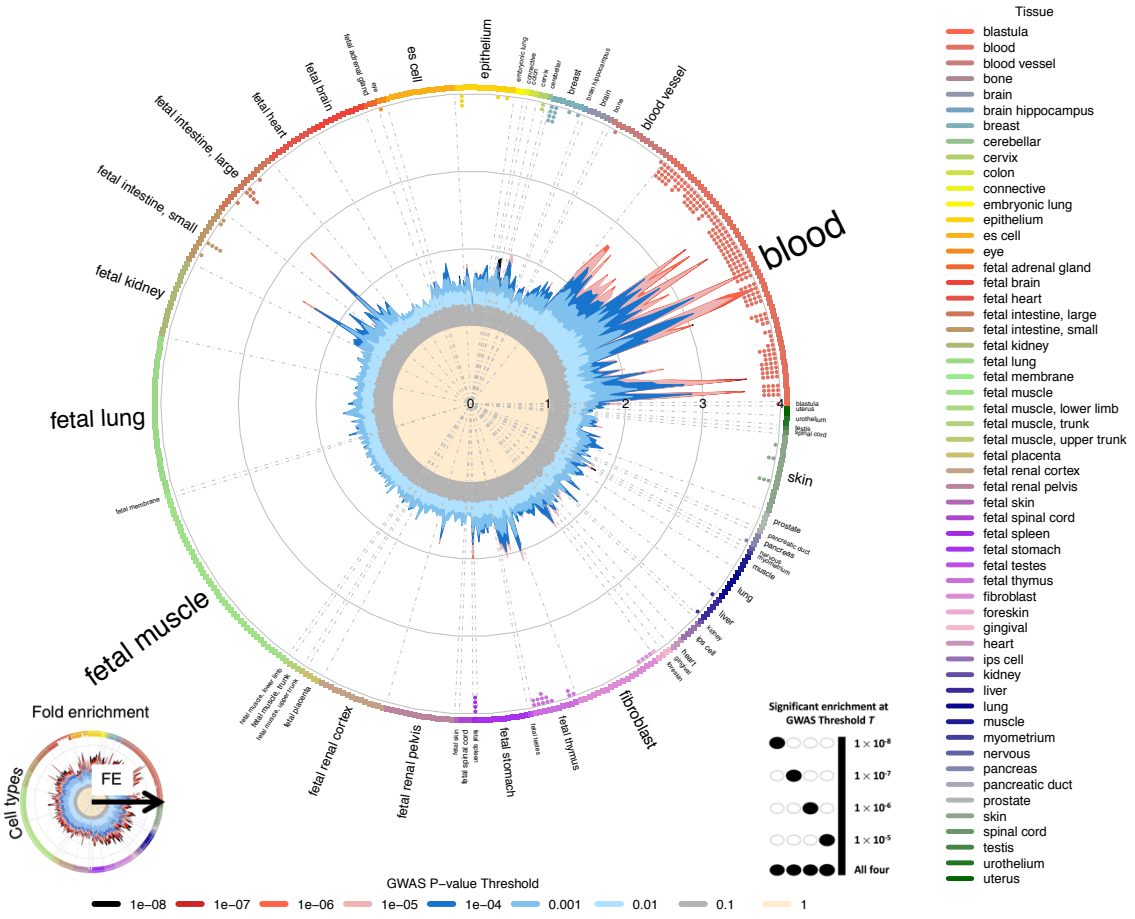


Figure 4- 2. GARFIELD functional enrichment analyses.

The wheel plot displays fold of enrichment in DHS hotspots for SLE GWAS statistics derived from the main SLE European cohort. The radial axis shows fold of enrichment (FE) calculated at each of eight GWAS P-value thresholds ($P < 1\times10^{-1}$ to $P < 1\times10^{-8}$) for all 424 annotations. Cell types are sorted by tissues, represented along the outside edge of the plot with font size proportional to the number of cell types from that tissue. Fold of enrichment values at different thresholds are plotted with different colours inside the plot, with colour legends shown at the bottom. Boxes and dots next to the tissue labels are coloured with respect to tissue, with colour legends shown on the right side. Dots along the inside edge of the plot (legend at the right bottom to the big wheel plot) denote whether the enrichment is significant (if there is a dot present) or not (no dot) for a given cell type at four GWAS P value thresholds, from $P < 1\times10^{-5}$ (outermost dot) to $P < 1\times10^{-8}$ (innermost dot).

4.3 SLE risk loci that are enriched in target tissue/cell types

According to the functional elements enrichment results from GARFIELD analysis, the functions of SLE risk loci are mainly enriched in immune cells, fetal intestine, spleen and breast. In order to fully exploit information from the epigenomics analysis, I further explored what risk loci are

contributing to the overwhelming representation in these cell types and tissues. I used GoShifter [146] to search for the loci that contributed most to the enrichment. DHS epigenomes for the target cell types (Monocytes, NK cells, B cells, T cells, and LCL) and tissues (fetal large intestine, fetal small intestine, spleen, and breast) were extracted from (http://egg2.wustl.edu/roadmap/data/byFileType/peaks/consolidated/narrowPeak/?C=N;O=D) (Table 4-4) for GoShifter analysis.

The locus score, indicating whether the tag SLE risk SNP (or SNPs in strong LD with the tagSNP, $r^2 \geq 0.8$) overlapped with the functional annotations being tested, for all SLE risk loci in each of the nine cell types/ tissues are shown in Table 4-5. The lower the locus score, the better the co-localization of target SNP and the functional annotation. By clustering the locus score according to cell types/tissues, we can see that some of the cell types are closer than others (see Figure 4-3). For example, monocytes and NK cells are in the same cluster, which are both mainly involved in the innate immune system. B cells are similar to LCL, as LCL is a B cell line transformed with EBV, leading to their commonalities. In addition, fetal small intestine and fetal large intestine are in the same cluster and the loci that contribute to the significant enrichment are all overlapped with the ones enriched in blood (including NK cells, monocytes, B cells and T cells) and showed preference in T cells. Almost one third of the loci overlapped with functional elements generally across a variety of cell types and tissues. For example, *UBE2L3* is significantly overlapped with DHS in all cell types and tissues. But some of the loci are preferably in specific tissue, e.g. *TERT*, a gene involved in the Estrogen-Dependent Breast Cancer Signalling pathway, is overlapped with DHS in breast specifically.

Table 4- 4. DNase-sequencing peaks (bed files) from Roadmap epigenomes.

RoadmapIndex	Celltype/Tissue	GROUP	STD_NAME	TYPE	ETHNICITY
E028	Breat vHEMC	Epithelial	Breast variant Human Mammary Epithelial Cells (vHMEC)	PrimaryCulture	Caucasian
E029	CD14 Monocytes	HSC & B-cell	Primary monocytes from peripheral blood	PrimaryCell	Hispanic, Unknown
E032	CD19 B cells	HSC & B-cell	Primary B cells from peripheral blood	PrimaryCell	Caucasian, Hispanic
E034	CD3 T cells	Blood & T-cell	Primary T cells from peripheral blood	PrimaryCell	Caucasian, Hispanic
E046	CD56 NK cells	HSC & B-cell	Primary Natural Killer cells from peripheral blood	PrimaryCell	Caucasian, Hispanic
E084	fetal intestine large	Digestive	Fetal Intestine Large	PrimaryTissue	Unknown
E085	fetal intestine small	Digestive	Fetal Intestine Small	PrimaryTissue	Unknown
E113	spleen	Other	Spleen	PrimaryTissue	Caucasian
E116	LCL	ENCODE2012	GM12878 Lymphoblastoid Cells	PrimaryCulture	Caucasian

Table 4- 5. A summary 'Locus Score' of SLE loci that are enriched in target cell types/ tissues.

targetSNP	chr	LocusGenes	Monocytes	B cells	T cells	NK cells	LCL	Breastv HEMC	Fetal intestine large	Fetal intestine small	Spleen
rs2476601	1	PTPN22	NA	NA	NA	NA	NA	0.032	NA	NA	NA
rs1801274	1	FCGR2A_FCGR2B	0.441	NA	NA	NA	NA	NA	NA	NA	NA
rs10911628	1	NCF2_EDEM3	0.339	0.322	NA	NA	NA	NA	0.163	NA	NA
rs34889541	1	PTPRC	0.409	NA	0.794	0.699	NA	NA	NA	NA	0.587
rs3024505	1	IKBKE_IL10	0.349	0.482	0.234	0.299	0.472	0.088	0.095	0.686	0.31
rs9782955	1	LYST	0.44	0.227	0.429	NA	0.322	0.22	0.168	0.202	NA
rs7579944	2	LBH	NA	0.547	0.721	NA	0.538	NA	0.404	0.433	0.572
rs6740462	2	SPRED2	NA	NA	NA	NA	NA	0.186	NA	NA	NA
rs6705628	2	DGUOK_CCT7	0.352	NA	0.341	NA	0.451	0.28	0.369	0.208	NA
rs3821236	2	STAT4	NA	NA	NA	0.467	NA	0.163	NA	NA	NA
rs3768792	2	IKZF2	NA	0.133	NA	NA	0.889	NA	NA	NA	NA
rs6445972	3	ABHD6_PXK	NA	0.813	NA	NA	NA	NA	NA	NA	0.815
rs2222631	3	GSK3B	NA	NA	NA	NA	0.249	NA	NA	NA	NA
rs564799	3	IL12A	0.725	0.6	NA	0.447	0.778	0.684	NA	NA	0.28
rs10936599	3	MYNN	0.687	0.596	0.549	NA	0.515	0.52	0.317	0.54	0.338
rs6762714	3	LPP_TPRG1_AS1	NA	0.123	NA	NA	0.137	NA	0.076	0.066	NA
rs4690229	4	DGKQ_IDUA	0.387	NA	NA	0.234	NA	NA	0.357	NA	NA
rs340630	4	PTPN13	NA	NA	NA	NA	0.796	NA	0.679	0.429	0.738
rs10028805	4	BANK1	NA	NA	NA	NA	NA	NA	NA	NA	0.329
rs907715	4	IL21	NA	NA	0.143	NA	NA	0.189	NA	NA	NA
rs7726159	5	TERT	NA	NA	NA	NA	NA	0.204	NA	NA	NA
rs7726414	5	TCF7	NA	NA	0.639	NA	NA	NA	NA	NA	NA
rs7708392	5	ANXA6_SLC36A1	0.542	0.377	0.644	NA	0.63	NA	NA	0.542	NA
rs2431697	5	PTTG1_UBLCP1	NA	NA	NA	NA	0.186	NA	NA	NA	NA
rs36014129	6	SLC17A1_BTN3A2	NA	0.484	NA	0.648	NA	NA	0.832	NA	NA
rs597325	6	BACH2	0.233	NA	0.714	0.357	NA	NA	NA	NA	NA
rs6568431	6	PRDM1_ATG5	NA	0.315	0.489	0.491	0.717	NA	NA	NA	NA
rs6932056	6	IL22RA2_PERP	0.962	0.848	0.986	0.948	0.976	0.938	0.587	NA	0.642
rs849142	7	HOXA9	NA	NA	NA	NA	NA	0.53	NA	NA	NA
rs4917014	7	IKZF1	0.383	0.222	0.699	0.495	0.344	NA	NA	NA	0.266
rs73135369	7	CCL24_HIP1_NCF1	NA	NA	NA	NA	NA	0.234	NA	NA	NA
rs4728142	7	TNPO3	0.708	NA	0.888	NA	NA	NA	NA	NA	NA
rs2736340	8	BLK_C8ORF13	0.723	0.873	NA	0.887	NA	0.734	NA	NA	0.538
rs7829816	8	RPS20_PLAG1	NA	NA	NA	NA	0.721	NA	NA	NA	NA
rs1966115	8	PKIA_ZC2HC1A	0.47	0.536	0.4	NA	0.844	NA	NA	NA	NA
rs877819	10	WDFY4	NA	0.247	NA	NA	0.307	NA	NA	NA	NA
rs4948496	10	ARID5B	0.825	0.918	0.92	0.685	0.876	0.709	0.945	0.86	0.848
rs4917385	10	USMG5	NA	NA	NA	NA	0.25	NA	NA	NA	NA
rs12802200	11	TMEM80_PHRF1	NA	0.938	0.932	0.709	NA	NA	NA	NA	NA
rs2732552	11	CD44	0.967	0.729	0.994	0.83	0.906	0.551	NA	NA	0.531
rs494003	11	RNASEH2C_CTSW	0.405	0.483	0.429	0.354	NA	0.486	0.365	0.37	NA
rs3794060	11	PPFIA1	0.715	0.548	0.616	0.672	0.766	0.645	0.778	0.594	NA
rs6590330	11	FLI1	NA	0.184	0.763	0.521	0.403	0.55	NA	NA	NA
rs34330	12	APOLD1_HTR7P	NA	NA	0.61	NA	NA	0.505	NA	NA	NA
rs4622329	12	NT5DC3	0.483	NA	NA	NA	0.265	NA	NA	NA	NA
rs10774625	12	ATXN2	NA	NA	0.422	NA	NA	NA	NA	NA	NA
rs1059312	12	SLC15A4	0.954	0.991	0.997	0.98	0.998	0.441	NA	NA	0.995
rs8016947	14	NFKBIA	NA	NA	NA	NA	NA	NA	NA	NA	0.237
rs2289583	15	CSK_ULK3	NA	NA	NA	0.464	0.576	0.414	NA	0.391	0.424
rs9652601	16	CLEC16A_SOCS1	0.456	0.869	0.883	NA	0.9	0.81	NA	NA	0.769
rs7197475	16	ITGAM_MAPK3	0.706	0.832	0.765	0.557	0.521	0.647	0.585	0.496	0.348
rs223881	16	CCL22	0.203	NA	NA	0.217	0.34	NA	NA	NA	NA
rs1170426	16	ZFP90	0.789	0.921	0.931	0.939	0.761	0.979	0.717	NA	0.424
rs2280381	16	IRF8	0.734	0.574	NA	0.909	NA	NA	NA	NA	0.249
rs2286672	17	RNF167	NA	NA	0.554	NA	NA	NA	NA	NA	NA
rs2941509	17	IKZF3	NA	1	1	1	1	0.907	0.982	0.729	1
rs930297	17	USP36_KIAA0195	0.495	0.56	0.462	0.296	0.314	0.162	0.286	0.3	NA
rs3093030	19	TYK2_ICAM1_ICAM4	0.991	0.995	0.995	0.995	0.986	0.821	0.765	0.793	NA
rs2305772	19	SIGLEC12_ZNF175	0.229	0.733	NA	0.384	0.404	NA	NA	NA	NA
rs4810485	20	SLC35C2_PLTP	0.519	0.665	0.215	0.344	NA	NA	NA	0.11	NA
rs11697848	20	RNF114	0.985	0.963	0.987	0.776	0.955	0.939	0.999	0.944	0.939
rs7444	22	UBE2L3	0.796	0.995	0.992	0.893	0.983	0.874	0.896	0.862	0.863

Note: NA is not available, which means that specific locus does not overlapped with the DHS hotspots in the according cell type. The number in this table for each SNP and each cell type is the 'locus score' obtained from GoShifter. Locus score is the probability that each locus overlaps an annotation by chance and defined as I_s/n, where I_s is the number of shifting iterations for which at least one SNP within an individual locus overlaps the annotation, and n is the total number of iterations. Loci with low scores are higher-priority candidates for further investigations, as they drive the significant enrichment observations. For more details, refer to 2.3.2. GoShifter.

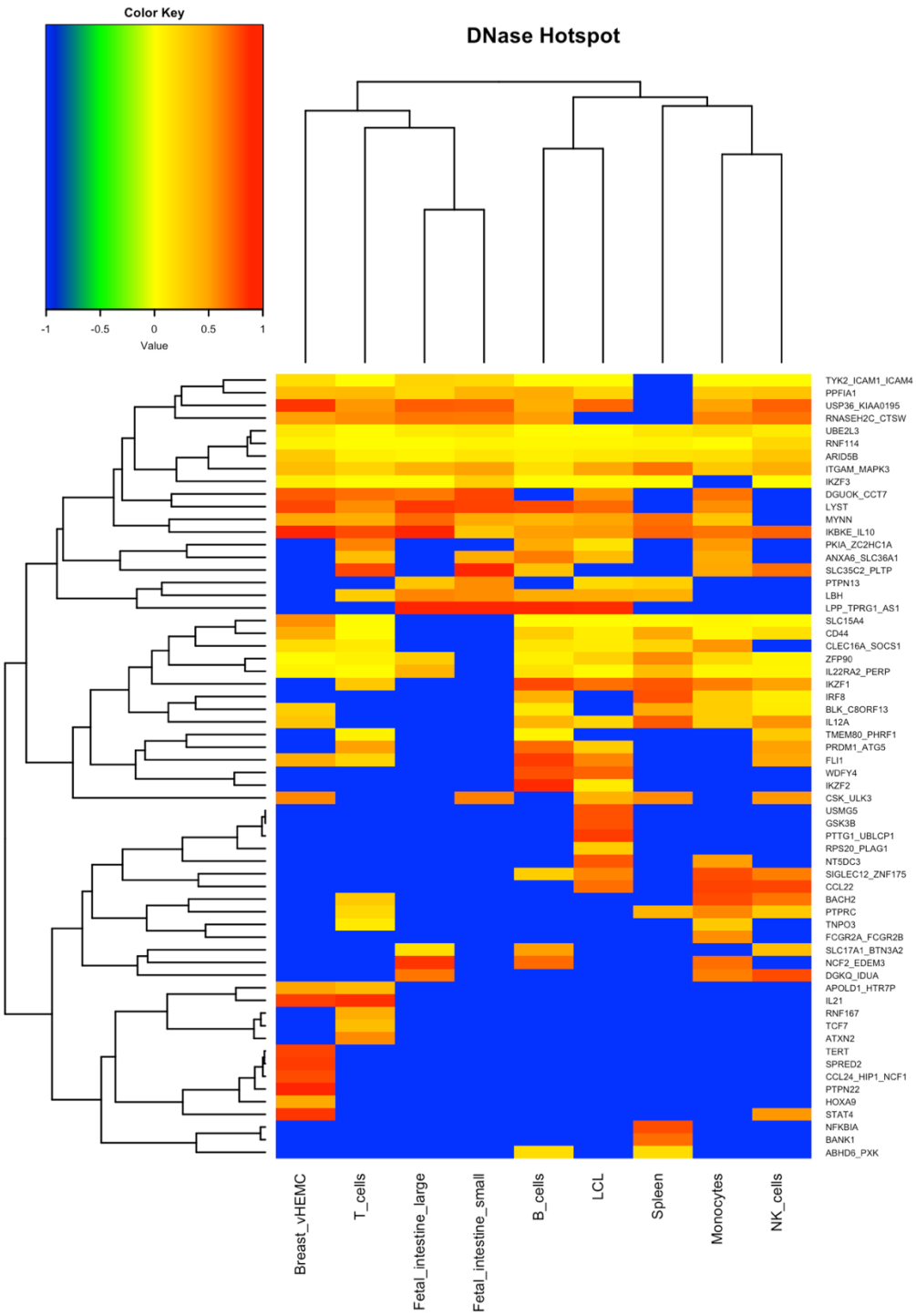


Figure 4- 3. SLE associated loci that contributing to the enriched cell types/tissues**.**
The 'pseudo' locus score derived from GoShifter is used to generate the heat map. By 'pseudo', I mean that not the 'real' locus score that shown in Table 4-5, but rather the '1-locus score' is applied. For loci that show no-overlapping with DHS (locus score is 'NA' in Table 4-5), '-1' is assigned as their 'pseudo' locus score to make sure all loci in all cell types have a 'pseudo' score for clustering. Each column represents a cell type/tissue and each row represents a locus (named according to the genes nearby). The redder, the better the overlap, while blue indicates no overlapping between the locus and the functional annotation.

4.4 Integrating eQTL and epigenetics to annotate SLE risk loci

Previous studies [96, 99] have shown that genetic, epigenetic and gene expression variance are strongly locally coordinated. Thus, I sought to identify shared genetic effects to link SLE risk genes to their putative regulatory elements. I considered all SLE risk alleles and asked whether the same variant / locus was associated with both gene expression and DHS hotspot. For this purpose, I simply compared if the SLE eSNP (SLE risk SNP found to be an eQTL) was overlapped with the DHS hotspot in which cell(s). By comparison, 66% (31/47) of the SLE eQTLs were shown to overlap with DHS hotspots (Table 4-6), denoting extensive local coordination of genetic influences on gene expression and DHS hotspots. For example, SLE risk allele – rs564799 is an eQTL that regulating the gene expression of *IL12A* in B cells, which is overlapped with the DHS peak locating around *IL12A* in not only B cells, but also monocytes, NK cells, and LCL (Figure 4-4), suggesting that the risk allele of SLE may regulate the expression of *IL12A* through the perturbation of DHS peak – the according chromatin accessibility.

Table 4- 6. Comparison of eQTL and DHS epigenetics in SLE risk loci.

targetSNP	chr	Locus.gene	eSNP?	eQTL in which cell(s)	DHS Hotspot ?	DHS overlapped in which Immune cells
rs4649203	1	IL28RA	✓	T cell	✗	
rs2476601	1	PTPN22	✓	LCL	✓	
rs1801274	1	FCGR2A_FCGR2B	✓	Monocytes	✓	Monocytes
rs17849502	1	NCF2_EDEM3	✓	LCL	✗	
rs10911628	1	NCF2_EDEM3	✗		✓	Monocytes + B cell
rs34889541	1	PTPRC	✗		✓	Monocytes + T cell + NK cell
rs2297550	1	IKBKE_IL10	✓	T cell + Monocytes	✗	
rs3024505	1	IKBKE_IL10	✓	Monocytes	✓	Monocytes + B cell + T cell + NK cell + LCL
rs9782955	1	LYST	✗		✓	Monocytes + B cell + T cell + LCL
rs7579944	2	LBH	✗		✓	B cell + T cell + LCL
rs13385731	2	RASGRP3	✓	LCL	✗	
rs6740462	2	SPRED2	✗		✓	
rs6705628	2	DGUOK_CCT7	✗		✓	Monocytes + T cell + LCL
rs2111485	2	IFIH1	✓	Monocytes	✗	
rs3821236	2	STAT4	✗		✓	NK cell
rs3768792	2	IKZF2	✗		✓	B cell + LCL
rs6445972	3	ABHD6_PXK	✓	T cell	✓	B cell
rs1132200	3	GSK3B	✓	LCL	✗	
rs2222631	3	GSK3B	✗		✓	LCL
rs564799	3	IL12A	✓	B cell	✓	Monocytes + B cell + NK cell + LCL
rs10936599	3	MYNN	✗		✓	Monocytes + B cell + T cell +LCL
rs6762714	3	LPP_TPRG1_AS1	✗		✓	B cell + LCL
rs4690229	4	DGKQ_IDUA	✓	LCL + WB + T cell + Monocytes + B cell	✓	Monocytes + NK cell
rs340630	4	PTPN13	✓	LCL	✓	LCL
rs10028805	4	BANK1	✗		✓	
rs907715	4	IL21	✗		✓	T cell
rs7726159	5	TERT	✗		✓	
rs7726414	5	TCF7	✗		✓	T cell
rs7708392	5	ANXA6_SLC36A1	✓	WB	✓	Monocytes + B cell + T cell + LCL
rs2431697	5	PTTG1_UBLCP1	✗		✓	LCL
rs36014129	6	SLC17A1_BTN3A2	✓	LCL	✓	B cell + NK cell
rs597325	6	BACH2	✗		✓	Monocytes + T cell + NK cell
rs6568431	6	PRDM1_ATG5	✗		✓	B cell + T cell + NK cell + LCL
rs2327832	6	IL22RA2_PERP	✓	LCL	✗	
rs6932056	6	IL22RA2_PERP	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs849142	7	HOXA9	✗		✓	
rs4917014	7	IKZF1	✓	WB	✓	Monocytes + B cell + T cell + NK cell + LCL
rs73135369	7	CCL24_HIP1_NCF1	✓	LCL	✓	
rs1167796	7	CCL24_HIP1_NCF1	✓	WB	✗	
rs4728142	7	TNPO3	✓	T cell + Monocytes	✓	Monocytes + T cell
rs2070197	7	TNPO3	✓	LCL	✗	
rs2980512	8	BLK_C8ORF13	✓	WB	✗	
rs2736340	8	BLK_C8ORF13	✓	LCL + WB + B cell	✓	Monocytes + B cell + NK cell
rs7829816	8	RPS20_PLAG1	✓	WB	✓	
rs1966115	8	PKIA_ZC2HC1A	✓	LCL + B cell + NK cell	✓	Monocytes + B cell + T cell + LCL
rs877819	10	WDFY4	✓	Monocytes	✓	B cell + LCL
rs4948496	10	ARID5B	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs4917385	10	USMG5	✓	LCL + WB	✓	LCL
rs12802200	11	TMEM80_PHRF1	✓	Monocytes	✓	B cell + T cell + NK cell
rs2732552	11	CD44	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs494003	11	RNASEH2C_CTSW	✓	Monocytes	✓	Monocytes + B cell + T cell + NK cell
rs3794060	11	PPFIA1	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs4639966	11	TMEM25_CD3D_C11ORF60	✓	WB	✗	
rs6590330	11	FLI1	✗		✓	B cell + T cell + NK cell + LCL
rs34330	12	APOLD1_HTR7P	✓	T cell	✓	T cell
rs4622329	12	NT5DC3	✗		✓	Monocytes + LCL
rs10774625	12	ATXN2	✓	Monocytes	✓	T cell
rs1059312	12	SLC15A4	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs8016947	14	NFKBIA	✗		✓	
rs11073328	15	SPRED1	✓	NK cell	✗	
rs2289583	15	CSK_ULK3	✓	WB + Monocytes	✓	NK cell + LCL
rs9652601	16	CLEC16A_SOCS1	✓	Monocytes	✓	Monocytes + B cell + T cell + LCL
rs16972959	16	PRKCB1	✓	LCL	✗	
rs7197475	16	ITGAM_MAPK3	✓	WB	✓	Monocytes + B cell + T cell + NK cell + LCL
rs34572943	16	ITGAM_MAPK3	✓	LCL	✗	
rs223881	16	CCL22	✗		✓	Monocytes + NK cell + LCL
rs1170426	16	ZFP90	✓	LCL + Monocytes	✓	Monocytes + B cell + T cell + NK cell + LCL
rs2280381	16	IRF8	✓	Monocytes	✓	Monocytes + B cell + NK cell
rs2286672	17	RNF167	✓	Monocytes + NK cell	✓	T cell
rs2941509	17	IKZF3	✓	LCL	✓	B cell + T cell + NK cell + LCL
rs930297	17	USP36_KIAA0195	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs3093030	19	TYK2_ICAM1_ICAM4	✓	LCL	✓	Monocytes + B cell + T cell + NK cell + LCL
rs2304256	19	TYK2_ICAM1_ICAM4	✓	WB + Monocytes	✗	
rs2305772	19	SIGLEC12_ZNF175	✓	LCL + Monocytes	✓	Monocytes + B cell + NK cell + LCL
rs4810485	20	SLC35C2_PLTP	✗		✓	Monocytes + B cell + T cell + NK cell
rs11697848	20	RNF114	✗		✓	Monocytes + B cell + T cell + NK cell + LCL
rs7444	22	UBE2L3	✓	LCL + WB + Monocytes	✓	Monocytes + B cell + T cell + NK cell + LCL
rs61616683	22	SYNGR1	✓	LCL + WB + Monocytes + B cell + NK cell	✗	

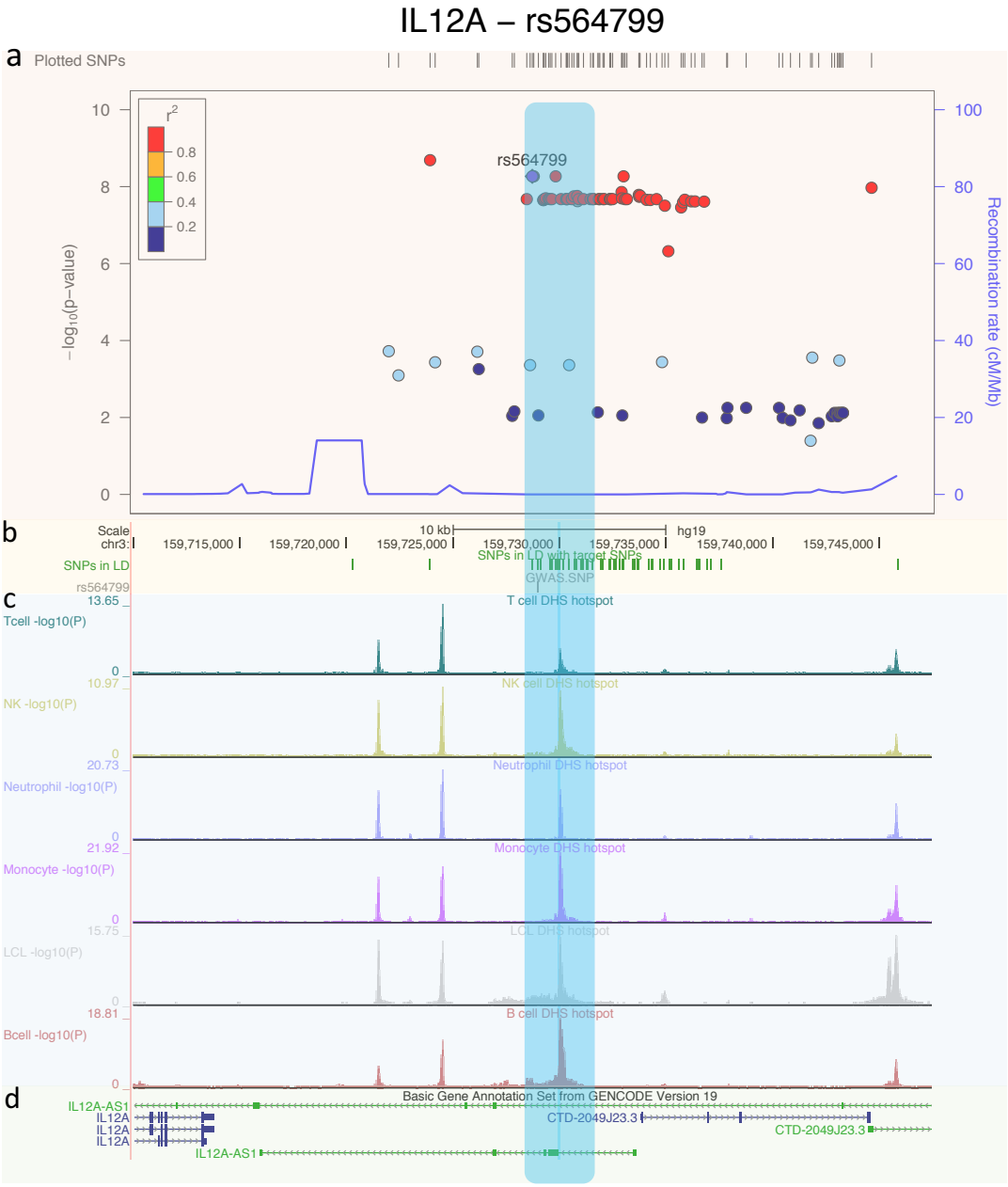


Figure 4- 4. An example of integrating eQTL and epigenetics to annotate SLE risk loci. a. The regional Manhattan plot by LocusZoom online tool (http://locuszoom.org) showed the –log10 P value of eQTL from the SNPs in the region of chr3: 159,720,000 – 159,745,000 versus gene expression of *IL12A* in ex vivo B cell. b. SNPs in strong LD ($r^2 \geq 0.8$) with the GWAS SNP rs564799. c. Signals of DHS peaks across different immune cell types (T cell, NK cell, Neutrophil, Monocyte, LCL and B cell). Data for the tracks were downloaded from Roadmap Epigenomes (http://egg2.wustl.edu/r oadmap/data/) and plotted using UCSC genome browser (http://genome-euro.ucsc.edu/). d. Basic Gene Annotation set from GENCODE version 19 of the target region.

4.5 Discussion

Large-scale GWASs analysed using the results from large-scale Epigenomic projects, such as ENCODE and Roadmap, make it possible to systematically investigate the functional effects of risk alleles mapped in regions associated with complex traits. They have greatly enhanced the annotation of putative functional consequences of non-coding variants in cell/tissue specific contexts which link to the underlying pathogenesis of diseases. However, current methods that aim to study SLE, a complex autoimmune disease involving many organs, are not robust due to the limited accessibility of the target organs from human beings. Most of the studies that aim to investigate the mechanism of lupus use model animals (often mice) as surrogates, though realising the major differences between human and mice [186, 187]. Here, I used a total of 424 genome-wide DHS hotspot annotations, derived from all the major organs and tissues of adults and fetus (Supplementary Table 2-1), incorporating genome-wide association summary results from the largest SLE cohort in European [43] to infer the regulatory role of SLE risk loci physically, i.e., in which tissue/s do the risk loci have their functions.

According to the SLE GWAS in the largest European cohort with a considerable sample size (N=10995) [187], most of the disease associated loci are located outside coding regions, leading to a question of how these loci contribute to the susceptibility of the disease. The genome is the assembly of DNA that makes each individual unique and the epigenome consists of proteins and chemical compounds that can attach to DNA and control the production of the genes in particular cells. DNA elements that used to be defined as 'junk' have revealed their functions through marking the genome and changing the way cells use the DNA's instructions [188, 189]. With the development of Epigenome Projects, rich resources of epigenome data are publicly available, facilitating improved understanding of non-coding DNA elements. I used epigenomic data derived from the ENCODE and Roadmap Epigenome Projects [96, 99, 189, 190] to prioritize SLE-relevant tissues and cell types by looking for the enrichment of GWAS signals within these functional genomic regions. Since DNaseI-seq identifies all genetic regulatory elements, including promoters, enhancers, silencers, insulators [191], and epigenomes generated by DNaseI-seq are available in almost all cell types and tissues, DHS hotspots serve as the best resource to search for the functional context for a target phenotype, in this case SLE.

GARFIELD is a non-parametric approach that leverages GWAS findings with regulatory or functional annotations to classify features relevant to a phenotype of interest. Both epigenomic data derived from ENCODE and Roadmap, and a comprehensive LD map generated by UK10K sequence data [145] were pre-processed and embedded as part of the GARFIELD package, making it very straightforward to obtain a global view of the functional context of SLE genetic associations. By applying GARFIELD, I found that SLE risk loci are enriched in regulatory elements in blood (B/T/NK/T cells, LCL), fetal intestine, and spleen, which are all critical in the innate and adaptive immune responses. However, no significant enrichment in kidney was observed, indicating indirect regulatory effects on target organs. A hypothesis is that risk alleles may assemble the circulating immune cells to target and accumulate in kidney, leading to lupus nephritis. Nevertheless, GARFIELD focuses on genome-wide enrichment analysis, not the details of specific loci, which does not allow us to infer what SLE risk loci are contributing to the enrichment within the functional genome in particular cell types. However, GoShifter is more adaptable and flexible enabling the user to infer functional disease associated loci and pinpoint putative causal variants. Therefore, I applied both methods to obtain a global view of SLE functional context and infer the risk loci contributing to the target tissue and cell types. Some SLE associated loci (Figure 4-3) were found to be enriched in DHS in immune cells, as well as in breast and intestine.

I realized that applying DHS only to infer the functional context of SLE may miss some regulatory elements. With the growing number of epigenomes becoming available, it would be better to integrate data from other resources, such like histone modifications and DNA methylation, to better understand the non-coding genome. However, epigenomes derived from ChIP-seq (such as H3K4me1, H3K4me3, H3K9ac, and H3K27ac) were not available for all the cell types and tissues – certainly not as comprehensively as DHS. Therefore, it would be biased (for false negatives) to use these regulatory epigenomes for the inference of functional context. In this chapter, I used the DHS as the functional annotation, which included 424 epigenomes representing all the common cell types and tissue of human body, providing a comprehensive overview of what the potential functional context may be. However, the functional regulatory annotations (DHS hotspots) are mainly derived from common cell types, while some of the potential rare cell types that are thought to be involved in SLE have not been characterized, such as plasma cells [192]. With the development of single cell RNA-sequencing (scRNA-seq) [193, 194], the resolution of identifying cellular differences has been largely extended, providing a better

understanding of the functions of rare cell populations. Hence, it will be a necessity to perform scRNA-seq in target tissue and cell types to fully fine map the functional context.

To the best of my knowledge, this is the first study of the functional context of SLE by integrating the epigenome and genome comprehensively and inferred the enrichment of SLE risk loci statistically. Though SLE is a complex disease with multiple organs involved, the overall regulatory pattern of SLE risk loci showed a significant enrichment in innate and adaptive immune cells, as well as in spleen, breast and fetal intestine, suggesting that organs manifested in SLE may not be targeted directly to trigger their impairment. Moreover, by using GoShifter, I have found specific loci contributing to the enrichment of specific cell types and tissues, providing evidence for further functional study. Taken together, I have increased our understanding of GWAS findings and provided insights into the role of immune cells in dissecting the underlying mechanism or pathogenesis of the disease.

Chapter 5. Use polygenic risk score and age onset of SLE for the inference of lupus nephritis

Chapter Summary

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease with a wide range of signs and symptoms among individuals and can involve many organs and systems. In particular, kidney involvement is the most common sub-phenotype, affecting 30-40% of SLE patient of European descent and accounting for significant morbidity and mortality. Genome-wide association studies (GWAS) have identified more than 80 loci that are associated with SLE. However, whether any of these genetic factors affect the sub-phenotype of SLE, i.e. lupus nephritis, is unknown.

I performed independent GWAS and a meta-analysis of two European cohorts to test for genetic association with the development of lupus nephritis (LN). To accomplish this, I conducted an in-case GWAS comparing the genetic data of those with renal disease, LN+ ($N_{cohort1}$ = 1152, $N_{cohort2}$ = 146) and those without renal disease, LN- ($N_{cohort1}$ = 1949; $N_{cohort2}$ = 378). The LN GWAS results were then combined with epigenetics data from ENCODE and Roadmap to study the enrichment patterns of LN risk variants. I found no significant renal associated loci or enrichment of renal associations in DNase I hypersensitive (DHS) hotspots.

I then calculated the genetic risk score (GRS) using SNPs reported to be significantly associated with SLE and found a significant correlation between the GRS and renal involvement and age of SLE onset in both European cohorts – the higher the GRS, the higher probability of renal disease ($P_{cohort1} = 7.2 \times 10^{-7}$; $P_{cohort2} = 0.00056$) and younger age of SLE onset ($P_{cohort1} = 2.4 \times 10^{-7}$; $P_{cohort2} = 0.021$).

In conclusion, I found that genetic risk factors may influence LN in a quantitative way (multiple genes' contribution). Therefore, I suggest that a GRS together with information on non-genetic risk factors will be useful for early diagnosis of LN patients in a clinical or research setting.

5.1 LN GWAS findings

LN occurs in approximately half of all SLE patients, and its frequency ranges from 25 to 75% depending on the population studied [195]. About one third of European SLE patients experience

renal disease [196]. Until recently, one of the most common cause of death in SLE patients was kidney failure. According to the lupus severity index (LSI) using the ACR criteria developed by Bello et al [197], renal involvement had the highest impact and particular strongly associated with disease severity, hence I chose LN as a representative sub-phenotype in this study.

The LN GWAS in the SLE main cohort, which comprised 1152 SLE patients with renal disease (LN+) and 1949 patients without renal disease (LN-), did not identify any genome-wide significant associated loci ($P \leqslant 5 \times 10^{-8}$) (Figure 5-1a). Consistently, no inflation (Genomic inflation factor: $\lambda = 1.014$) was observed in the QQ plot (Figure 5-1d). 97 SNPs (35 regions) were observed to have less stringent associations with a $P \leqslant 1 \times 10^{-5}$ (Supplementary Table 5-1). Similarly, none of the SNPs reached genome-wide significance ($P \leqslant 5 \times 10^{-8}$) in the SLEGEN cohort [41] ($\lambda = 1.023$) (Figure 5-1b and 5-1e) and 101 SNPs (23 regions) were observed to have a $P \leqslant 1 \times 10^{-5}$ (Supplementary Table 5-2). However, none of the SNPs with $P \leqslant 1 \times 10^{-5}$ in the main GWAS (Supplementary Table 5-1) replicated in the SLEGEN GWAS (Supplementary Table 5-2). In addition, there is no variant that passed genome-wide significance in the meta-analysis of the SLE main cohort and SLEGEN cohort for LN GWAS ($\lambda = 0.9565$) (Figure 5-1c and 5-1f) and 92 SNPs (21 regions) were identified to have less stringent associations with a $P \leqslant 1 \times 10^{-5}$ (Supplementary Table 5-3).

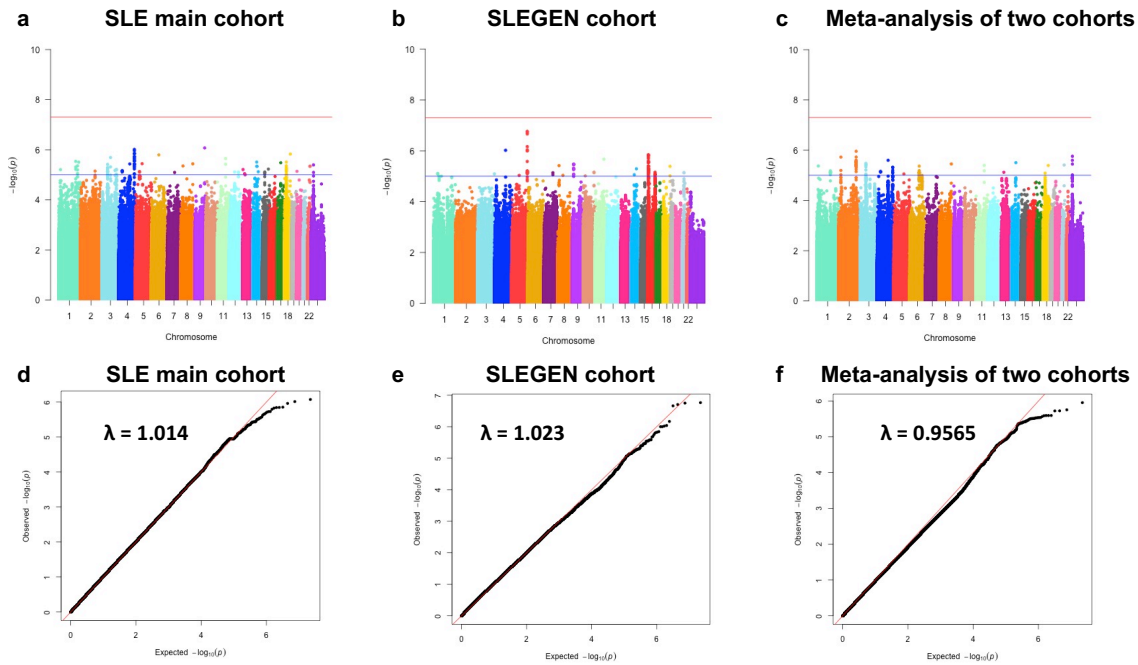


Figure 5- 1. Genome-wide scans of LN associated variants.

(Upper) Manhattan plots showing the -log10-transformed p values (y axis) against physical genomic position (x axis) for each SNP in (a) the SLE main cohort, (b) the SLEGEN cohort, and (c) the meta-analysis of these two cohorts. The red horizontal line represents the threshold for genome-wide significance ($P \leqslant 5 \times 10^{-8}$) and the blue horizontal line represents the threshold for suggestive significance ($P \leqslant 1 \times 10^{-5}$).

(Lower) Quantile-quantile plots showing the observed distribution of -log10-transformed p values (y axis) by the expected distribution (x axis) under the null hypothesis of no association (diagonal line) for (a) the SLE main cohort (genomic inflation factor, λ = 1.014), (b) the SLEGEN cohort (λ = 1.023), and (c) the meta-analysis of these two cohorts (λ = 0.9565).

5.2 Less stringent LN-associated variants are not enriched in DHS Hotspots

Most of the less stringent ($P \leqslant 1 \times 10^{-5}$) LN risk variants identified in this study are located outside exonic regions, where the underlying molecular mechanisms are poorly defined. To evaluate the functional properties of these variants, I estimated the extent to which these variants were non-randomly distributed across various coding, non-coding regulatory and cell-type-specific elements across the genome. 424 genome-wide data sets derived from the most widely studied annotation type, DNase I hypersensitivity site (DHS) hotspots, from the ENCODE and Roadmap projects were retrieved for analyses. A nonparametric approach, GARFIELD [143], was applied to derive fold of enrichment statistics for the suggestive SLE renal disease risk variants within each cell type/tissue where SNPs were selected on the basis of their strength of association (LN GWAS P value). Following the enrichment analysis, I found that none of the suggestive LN genetic associations, either from the SLE main cohort, the SLEGEN cohort, or the meta-analysis of these two cohorts, were enriched in DHS hotspots in any of the cell types or tissues being tested, neither in immune cells in the blood nor the target tissue (kidney) (Figure 5-2).

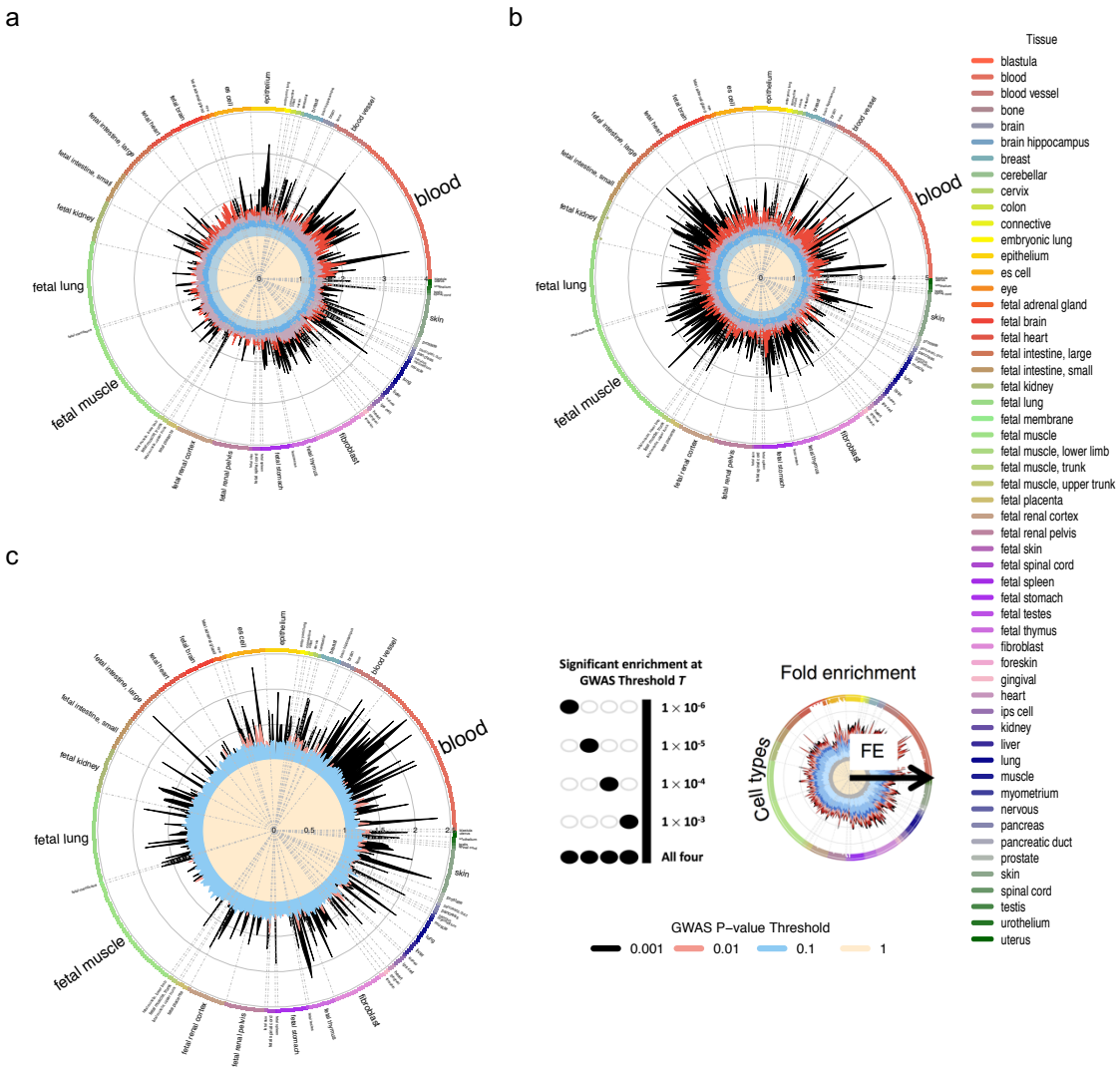


Figure 5- 2. GARFIELD functional enrichment analyses of LN GWAS statistics.

The wheel plot displays fold of enrichment in DHS hotspots for LN GWAS statistics of (a) the SLE main cohort, (b) the SLEGEN cohort, and (c) meta-analysis of these two cohorts. The radial axis shows fold enrichment (FE) calculated at each of six GWAS P-value thresholds: P < 1×10-1 to P < 1×10-6 for (a, b and c) for all 424 cell types. Cell types are sorted by tissues, represented along the outside edge of the plot with font size proportional to the number of cell types from that tissue. Fold enrichment values at different thresholds are plotted with different colours inside the plot, with colour legends shown at the bottom of each wheel plot. Boxes and dots next to the tissue labels are coloured with respect to tissue, with colour legends shown on the right side. Dots along the inside edge of the plot (legend at the right bottom to the wheel plot) denote whether the enrichment is significant (if there is a dot present) or not (no dot) for a given cell type at the GWAS P value thresholds, from the most significant P value (innermost dot) to the less significant P value (outermost dot).

5.3 Genetic risk loading of SLE is significantly correlated with LN

Due to the lack of significant associations for LN within SLE patients, I calculated a genetic risk score (GRS) to see if the accumulative burden of SLE risk alleles was correlated with LN. The GRS (see 2.5.1 GRS Calculation) were based on odds ratios reported for SLE susceptibility loci, which were summarized from the SLE GWAS studies [64]. The GRS in SLE patients and healthy controls for both independent cohorts from European population showed a clear higher GRS in SLE patients than healthy controls (Figure 5-3).

In terms of the GRS in patients with and without renal disease, a significantly higher GRS was observed in the group of patients with LN+ compared to patients with LN- (Figure 5-3). In the SLE main cohort, the mean (SD) of the GRS was 21.41 (2.06) for LN+ patients and 21.03 (2.12) for LN- patients ($P = 7.2 \times 10^{-7}$); the mean for the SLEGEN cohort was 21.66 (2.25) for LN+ patients and 20.89 (2.2) for LN- patients ($P = 5.6 \times 10^{-4}$). Moreover, I saw a significant increasing trend of GRS over levels of diseases: Healthy control, LN- patients, and LN+ patients, with a trend P-value = 1.10×10^{-270} in the SLE main cohort and a P-value = 6.55×10^{-48} in the SLEGEN cohort (Table 5-1).

Another way to show the differences in the distribution of GRS between SLE patients with and without LN is to partition the participants. I did this by defining risk stratum, starting at the mean of GRS in the SLE main cohort and grouping individuals who were within 0.25 SD of the mean as category 4. Individuals in category 4 can be regarded as those with the mean risk of renal disorder in SLE patients (Figure 5-4). Six subsequent categories (1-3 and 5-7) are defined by the increasing intervals of GRS. In the SLE main GWAS, individuals in category 7 (i.e., those with the largest GRS) had 1.20 times (95%CI: 0.92 -1.55) increased odds of developing renal disease compared with those in category 4 (Table 5-2). While patients in category 1 (i.e., those with the lowest GRS) were less likely to have LN compared with people in category 4 (OR = 0.58, p-value = 0.0012). When category 7 was compared with category 1, which contains the individuals with the lowest risk of renal involvement, there was a 2.05 (95% CI: 1.45 - 2.88) times increase in the odds of developing LN. The application of the GRS partitioning method in the SLEGEN cohort gave similar results: individuals in category 7 (i.e., those with the largest GRS) had a mean increase in odds of developing LN of 1.86 (0.92 - 3.77) compared with individuals in category 4. The risk increased to 2.99 (1.13 - 7.95) when category 7 was compared with category 1.

Table 5- 1. Summary of Genetic risk score GRS of the studied cohorts.

Cohorts	SLE VS Healthy controls			LN+ VS LN-		
	GRS * of SLE	GRS of Healthy controls	t test P-value	GRS of LN+	GRS of LN-	t test P-value
SLE main cohort	21.14 (2.1) n = 4036	19.22 (1.87) n = 6959	1.34E-302	21.41 (2.06) n = 1152	21.03 (2.12) n = 1949	7.22E-07
SLEGEN cohort	21.11 (2.25) n = 533	19.28 (1.92) n = 2543	7.57E-57	21.66 (2.25) n = 146	20.89 (2.24) n = 378	0.00056

Notes: GRS for SLE main cohort and SLEGEN cohort are generated by 95 non-MHC SNPs and 7 MHC SNPs, a total of 102 SNPs. All p-values were derived from two-tail *t*-test.

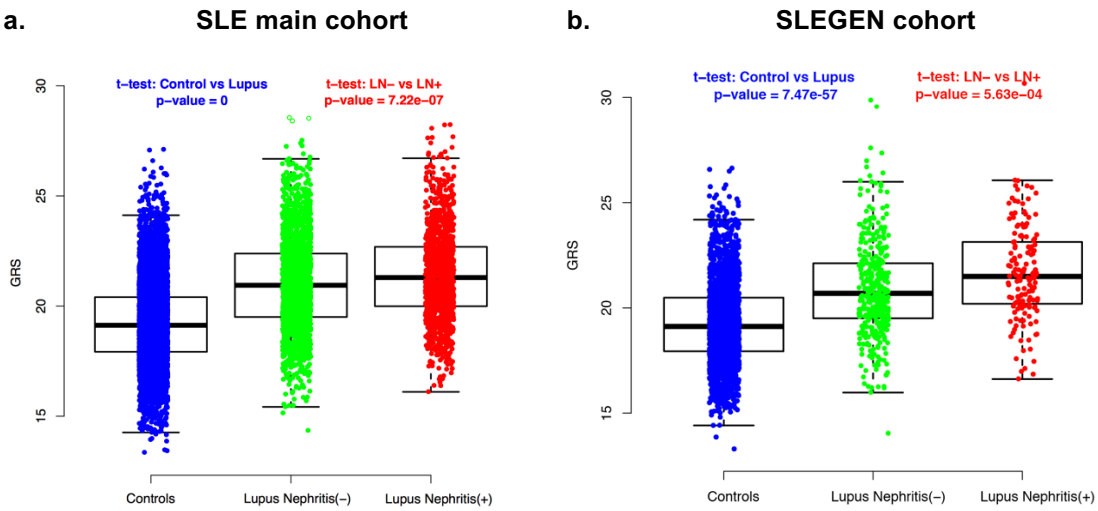


Figure 5- 3. GRS over levels of disease: Controls / Lupus Nephritis (-) / Lupus Nephritis (+).

The box-and-whisker plots show the summary GRS for each level of the disease in the SLE main cohort (a) and the SLEGEN cohort (b). The bottom line of the box is the 1st quantile, the top line is the 3rd quantile, and the box is divided at the median. The two whiskers outside the box extend to the smallest and largest observations within 1.5 times of interquartile range. On top of the boxplot is the strip chart, with each dot represent one individual.

Table 5- 2. GRS and odds ratios of SLE renal disease in SLE main cohort and SLEGEN cohort

	SLE main cohort			
	LN- (frequency, %)	LN+ (frequency, %)	OR (95% CI)	p-value
G1	198 (10.2)	69 (5.99)	0.58286 (0.42335 - 0.80248)	**0.0012**
G2	264 (13.5)	128 (11.1)	0.81094 (0.62041 - 1.06)	0.1424
G3	361 (18.5)	208 (18.1)	0.96369 (0.76018 - 1.2217)	0.8064
G4	378 (19.4)	226 (19.6)	1.0 (reference)	1.0000
G5	316 (16.2)	220 (19.1)	1.1644 (0.91753 - 1.4777)	0.2333
G6	209 (10.7)	142 (12.3)	1.1364 (0.86799 - 1.4878)	0.3891
G7	223 (11.4)	159 (13.8)	1.1925 (0.91781 - 1.5494)	0.2107
G1 VS G7 *	...	...	2.046 (1.4544 - 2.8782)	**0.000049**

	SLEGE Cohort			
	LN- (frequency, %)	LN+ (frequency, %)	OR (95% CI)	p-value
G1	31 (8.2)	7 (4.79)	0.62097 (0.24162 - 1.5959)	0.4418
G2	33 (8.73)	9 (6.16)	0.75 (0.31338 - 1.7949)	0.6661
G3	65 (17.2)	22 (15.1)	0.93077 (0.47519 - 1.8231)	0.9699
G4	66 (17.5)	24 (16.4)	1.0 (reference)	1.0000
G5	43 (11.4)	26 (17.8)	1.6628 (0.84679 - 3.2652)	0.1901
G6	33 (8.73)	15 (10.3)	1.25 (0.57956 - 2.696)	0.7106
G7	34 (8.99)	23 (15.8)	1.8603 (0.91837 - 3.7683)	0.1207
G1 VS G7 *	...	...	2.9958 (1.1288 - 7.9509)	**0.0426**

Notes: Data are in n(%) and OR(95% CI).

* OR is derived from the comparison of G1 VS G7. All others use G4 as a reference group.

** p-value is derived from the chi-square test.

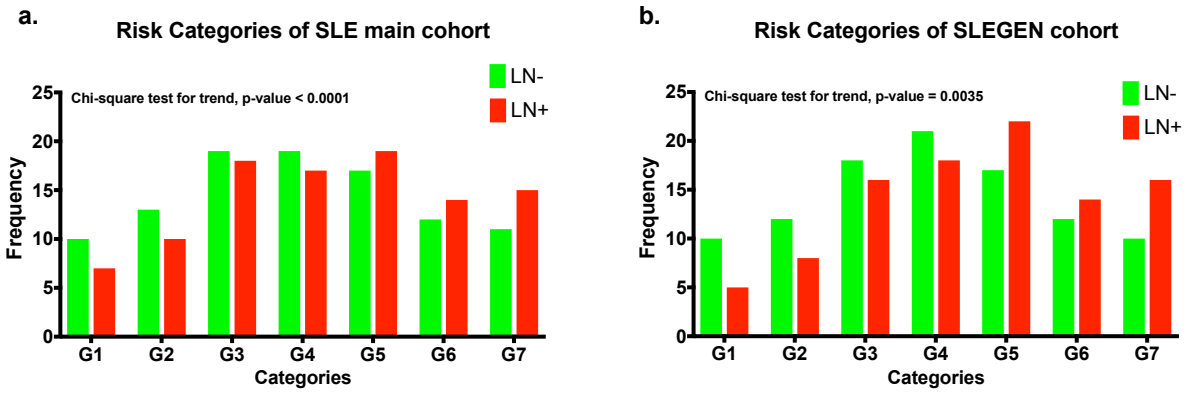


Figure 5- 4. Distribution of GRS by SLE Renal risk categories.

SLE renal risk categories for (a) the SLE main cohort and (b) the SLEGEN cohort. Seven categories of genetic risk are defined by the SLE main cohort, with G1 being the lowest risk category. The distribution of LN- patients are plotted as green while LN+ patients are plotted as red.

5.4 Age of SLE onset is significantly correlated with LN

In terms of the age of SLE onset in patients of LN- and LN+, a significant earlier age onset was observed in the LN+ patients compared to LN- patients (Figure 5-5). In the SLE main cohort, when testing the relationship between the age of SLE onset and LN in the SLE main cohort, the mean (SD) for age of SLE onset was 29 (12) for LN+ patients and 35 (13) for LN- patients ($P = 2.8 \times 10^{-27}$); the mean for the SLEGEN cohort were 28 (11) and 35 (13) for patients of LN+ and LN-, respectively ($P = 1.1 \times 10^{-8}$).

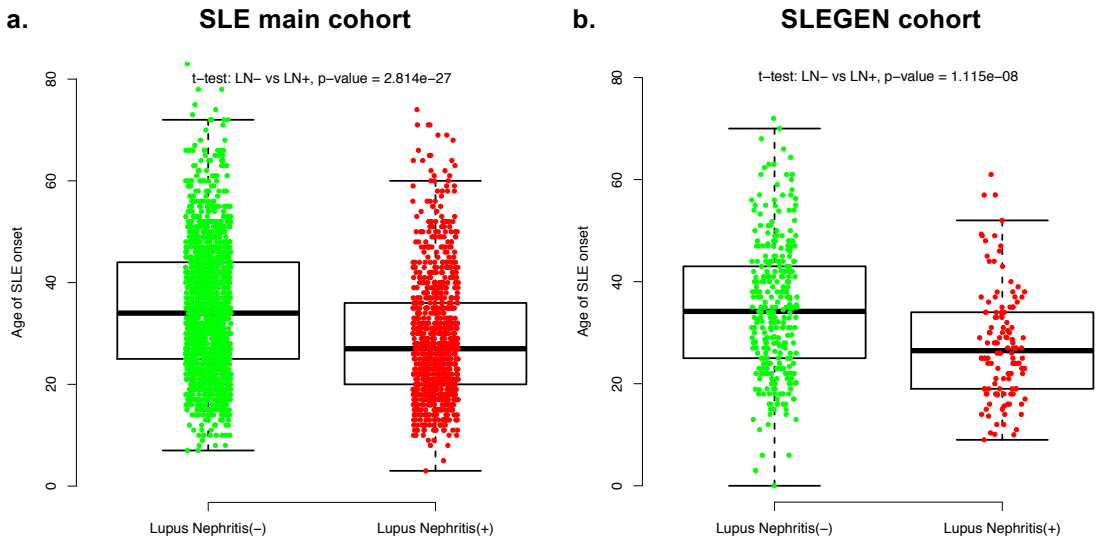


Figure 5- 5. Age of SLE onset in patients of Lupus Nephritis (-) / Lupus Nephritis (+). The box-and-whisker plots show the age of SLE onset for each level of the disease in the SLE main cohort (a) and the SLEGEN cohort (b). The bottom line of the box is the 1st quantile, the top line is the 3rd quantile, and the box is divided at the median. The two whiskers outside the box extend to the smallest and largest observations within 1.5 times of interquartile range. On top of the boxplot is the strip chart, with each dot represent one individual.

5.5 Use GRS and age onset of SLE to predict LN

When testing the association of GRS with age-of-onset in the SLE main cohort, a significant correlation can be seen – the higher the GRS, the younger age of SLE onset ($P = 2.4 \times 10^{-7}$) (Figure 5-6). This correlation can also be observed in the SLGEN cohort (P= 0.021). In the SLE main cohort, when testing the relationship of GRS with age-of-onset separately in the LN+ and LN- patients, the GRS was negatively correlated with the age-of-onset in both LN+ group ($P = 7.1 \times 10^{-5}$) and LN- group (P = 0.05). In the SLEGEN cohort, the correlation of GRS was also inversely correlated with age-of-onset in the LN- group (p = 0.011), while in the LN+ group, no significant correlation was observed (P = 0.28).

By adding the age-of-onset into the model for discriminating LN+ and LN- patients, the discrimination was significantly enhanced – the differences in AUC (the area under the receiver operating characteristic curve, a measure of the ability of a model to discriminate between LN+ and LN- patients) are statistically significant in both cohorts, see Table 5-6. In the SLE main cohort, AUC for the genetic-only model was 0.596 and for the genetic plus age-of-onset model was 0.665. In the SLEGEN cohort, AUC for the genetic-only model was 0.594; adding age-of-onset to the model increased the AUC to 0.682. The level of predictability remains modest, but our model clearly and consistently performs better than by chance (Figure 5-7).

Table 5- 3. AUC for ROC curves and comparisons between models to predict susceptibility to SLE patients with renal involvement.

Models	SLE main cohort		SLEGEN cohort	
	AUC	p value	AUC	p value
Model 1: covariates only	0.573	..	0.553	..
Model 2: GRS with covariates	0.596	0.000601*	0.594	0.0374*
Model 3: GRS with covariates and age-of-onset	0.665	1.46×10^{-9}**	0.682	0.00148**

*p value was generated from bootstrap test for two ROC curves (Model 1 VS Model 2); ** p value was generated from bootstrap test for two ROC curves (Model 2 VS Model 3). PCAs are covariates derived from the GWAS of each cohort. For the SLE main cohorts, the covariates included sex and four principle components derived from its original GWAS analysis. For the SLEGEN cohort, all the SLE patients are female, only three principle components derived from its corresponding GWAS study were included. ROC=receiver operating characteristic. AUC=area under the ROC curve.

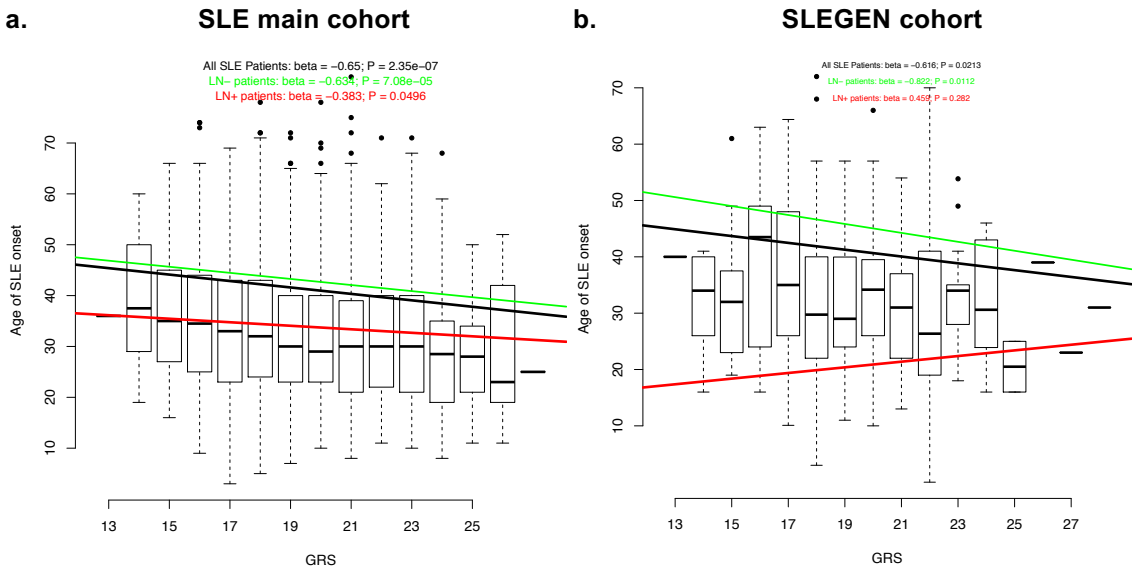


Figure 5- 6. Association of GRS and Age of SLE onset.

Plots of association of GRS and age of onset for (a) the SLE main cohort, and (b) the SLEGEN cohort. GRS was rounded with two significant digits, which divided all samples into a serial of groups. The box-and-whisker plots and linear regression lines (black for all SLE patients, green for LN- patients, Red for LN+ patients) show the Age of SLE onset (y axis) against allGRS (x-axis).

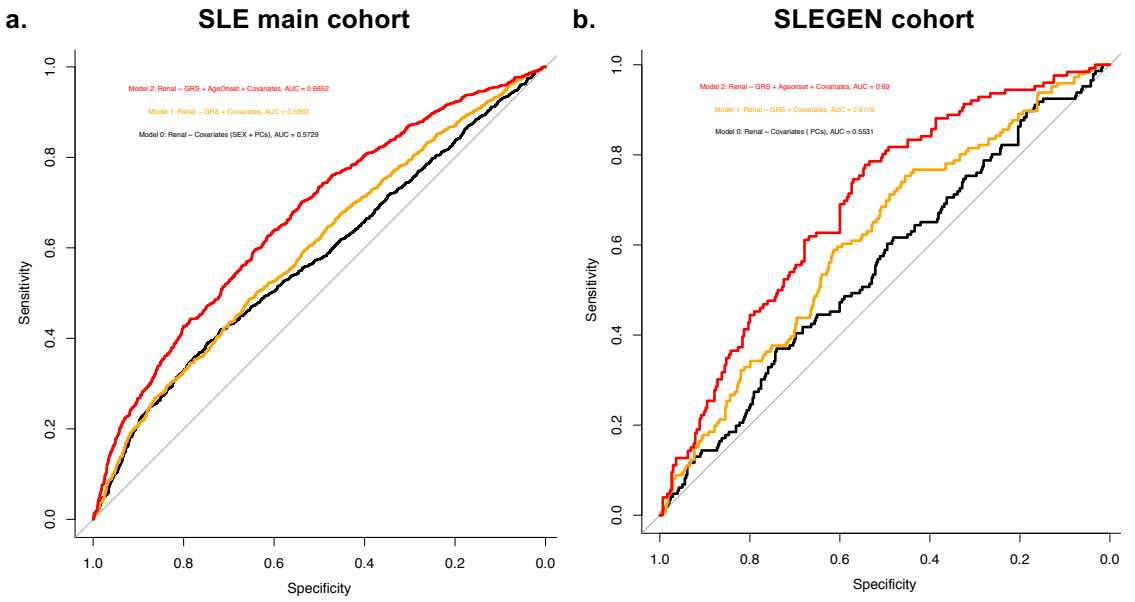


Figure 5- 7. Receiver Operating Characteristic (ROC) Curves for models predicting a diagnosis of LN in European cohorts.

In the SLE main Cohort (a), the results for these separate models to predict patients of LN are plotted: Model 0: without GRS – only covariates are included (black); Model 1: GRS and covariates (orange); Model 2: GRS, Age of Onset and covariates (red). In the SLEGEN Cohort (b), we repeated the analysis shown in (a). The same presentation scheme is used. The participants are all female. PCAs are covariates derived from the GWAS of each cohort. AUC=area under the ROC curve.

5.6 Discussion

Until recently, the most common cause of death in SLE patients was kidney failure. Though with better therapies (e.g. dialysis and kidney transplantation), the frequency of death from kidney disease has decreased sharply, kidney failure is still potentially fatal in some people with SLE and causes significant morbidity. According to the lupus severity index (LSI) using the ACR criteria developed by Bello et al [197], renal involvement had the highest impact and particularly more strongly associated with disease severity, hence I used renal involvement as a proxy of SLE severity in my study.

In the SLE within-case renal GWAS studies, I observed no genome-wide significant signals in either the SLE main cohort or the SLEGEN cohort, or meta-analysis of these two. Both datasets had genetic variants with less stringent P values ($P \leqslant 1 \times 10^{-5}$) for renal association, but none of them were replicated in the other cohort. Considering the sample size of both cohorts are relatively small, I applied an online genetic power calculator (http://zzz.bwh.harvard.edu/gpc/) to calculate the power of our current sample size for the GWAS study (Table 5-4). If we assume the effect size of SLE renal risk allele is similar to the ones that seen in SLE GWAS, we expected that the odds ratio (OR) of the risk allele would be majorly between 1.0 and 2.0. Therefore, I calculated the genetic power under a variety of parameters, including OR, risk allele frequency (RAF) and alpha. As showed in Table 5-4, we have a power of $\geqslant 0.8$ to detect a genetic risk variant who has an OR = 1.2 and RAF = 0.5 or an OR = 1.3 and RAF = 0.1 when alpha = 5×10^{-8}. When the OR of the risk allele is $\geqslant 1.4$, we have sufficient power to detect the genetic association of the common variants (RAF $\geqslant 0.05$) at the GWAS significant threshold of $P = 5 \times 10^{-8}$.

Due to the lack of significant genetic associations in SLE Renal GWAS, I then tested the hypothesis that the genetic risk loading of SLE may correlate with kidney involvement in a quantitative way – multiple genes with modest effect size. Therefore, a genetic risk score (GRS) summing up all the SLE risk is calculated. In both European cohorts, the SLE main cohort and the SLEGEN cohort, the GRS is significantly higher in patients with renal disease than patients without. In addition, patients with a higher GRS are more likely to have the renal involvement at younger age, indicating the strong genetic background of SLE development. Patients with late onset are more likely to have lower GRS, as the onset of SLE at older age is more likely to be triggered by the cumulative deleterious environmental factors, such as UV, cigarettes, and virus.

These findings provide more evidence to support the opinion that younger-age onset lupus is generally more severe than older-onset lupus as reported previously [198-200].

One may argue that if the severity of SLE is driven by multiple genes' contribution in a quantitative way, the more risk alleles that are added to the model, then the better the model would fit. In this study, I show that a GRS is a useful tool for the classification of SLE renal+ and SLE renal- groups. The renal association P values of the 102 SNPs (of 78 SLE risk loci) in the SLE main cohort and the SLEGEN cohort are strongly inflated as show in the QQ plots (Figure 5-8), suggesting the cumulative genetic burden from multiple SLE risk genes with modest effect.

In conclusion, I found that the genetic risk factors that influence the severity of SLE are through a quantitative way (multiple genes' contribution), but not a qualitative way (no significant associations specific to SLE renal were identified in either SLE renal GWASs). The inclusion of SLE risk alleles into a GRS can modestly predict the severity of SLE patients, shows consistent discriminatory ability in independent cohorts, and is enhanced by the inclusion of age-of-onset into the model. Therefore, I suggest that the GRS will make a contribution to predicting renal disease in SLE patients in a clinical or research setting.

Table 5- 4. Genetic power of current sample size under different parameters.

Genetic power	Alpha = 1e-5					Alpha = 1e-6					Alpha = 5e-8				
RAF	0.05	0.1	0.2	0.5	0.8	0.05	0.1	0.2	0.5	0.8	0.05	0.1	0.2	0.5	0.8
OR = 1.1	0.00	0.02	0.07	0.20	0.05	0.00	0.00	0.03	0.10	0.02	0.00	0.00	0.01	0.03	0.00
OR = 1.2	0.14	0.54	**0.93**	**0.99**	0.76	0.06	0.36	**0.84**	**0.97**	0.59	0.02	0.18	0.67	**0.92**	0.37
OR = 1.3	0.73	**0.99**	**1.00**	**1.00**	**1.00**	0.56	**0.97**	**1.00**	**1.00**	**0.99**	0.34	**0.91**	**1.00**	**1.00**	**0.95**
OR = 1.4	**0.99**	**1.00**	**1.00**	**1.00**	**1.00**	**0.97**	**1.00**	**1.00**	**1.00**	**1.00**	**0.91**	**1.00**	**1.00**	**1.00**	**1.00**
OR = 1.5	**0.89**	**1.00**	**1.00**	**1.00**	**1.00**	**0.78**	**1.00**	**1.00**	**1.00**	**1.00**	**0.58**	**1.00**	**1.00**	**1.00**	**1.00**
OR = 2.0	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**

*This table showed the genetic power for the sample size of 1100 (SLE Renal+) and 1980 (SLE Renal-) under a variety of ORs and RAFs. Text in red indicates that we have a power of ≥ 0.8 to detect the genetic associations under corresponding alpha, RAFs and OR. RAF = risk allele frequency; OR = odds ratio.

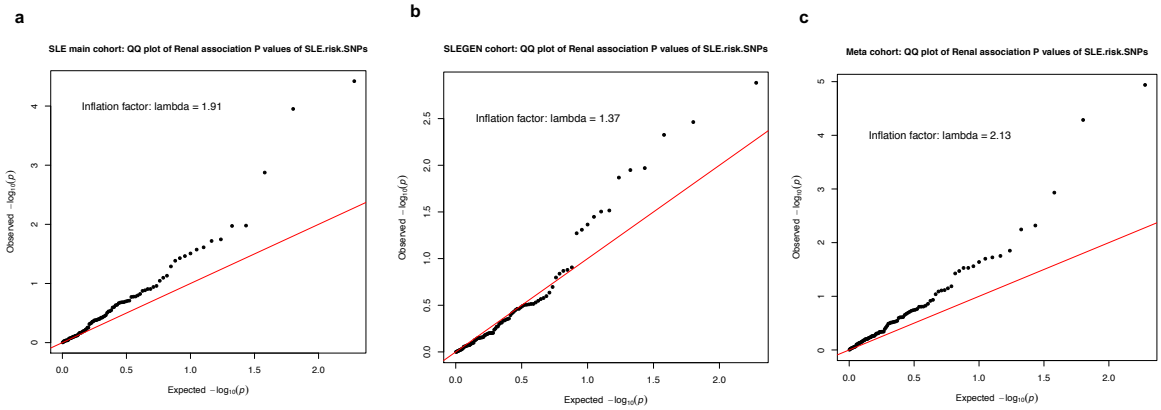


Figure 5- 8. Quantile-quantile plots of LN GWAS results.

QQ plots showed the observed distribution of -log10-transformed p values (y axis) by the expected distribution (x axis) under the null hypothesis of no association (diagonal line) for the SLE main cohort (a), the SLEGEN cohort (b) and the meta-analysis of these two cohorts (c). The p-values for the QQ plots were derived from SLE renal association test of the 77 SLE risk loci (non-MHC region) which used for the GRS calculation.

Chapter 6. Functional studies of the Ikaros family and OX40L using Flow-cytometry

Chapter Summary

Previous studies on the Ikaros transcription factor family and TNF superfamily gene TNFSF4 have shown that the genes encoding these proteins are candidate susceptibility genes in SLE and other autoimmune diseases. Nevertheless, how these factors affect the pathogenesis of SLE remains largely unknown. By measuring the protein level of OX40L, IKZF1, IKZF2, and IKZF3 on the immune subpopulations in PBMC derived from 22 SLE patients and 20 healthy controls, I aim to tease out the contributions of specific cells for these molecular factors under disease status, providing clues for further experimental study.

This study revealed, for the first time, the protein expression pattern of OX40L and Ikaros transcription factor family (IKZF1/2/3) across a range of immune cells, including T cells, B cells, monocytes, NK cells, dendritic cells, as well as their sub-populations. Through the cell proportions analysis, I showed that cell proportions of immune cells derived from the adaptive immune system, such as B cells and activated CD4+ T cells, were increased, while innate immune cells like DCs and NK cell were decreased in SLE patients comparing to healthy controls. In general, I observed a trend of increase for the protein level of OX40L and a trend of decrease for IKZF1, IKZF2 and IKZF3 in SLE patients, with some differences between cases and controls being statistically significant. Specifically, OX40L on B cell subsets was significantly increased (p-value ≤ 0.05) in SLE patients, suggesting OX40L may be involved in the pathogenesis of SLE though its role in B cells. Moreover, I showed that both IKZF1 and IKZF3 were significantly decreased in some of the activated CD4 T cells (p-value ≤ 0.05) in SLE patients, such as activated Th2 cells, while no significant differences between cases and controls were observed for IKZF2. Finally, by analysing the correlation between protein levels of target molecules with SLEDAI, I found that OX40L on Th17 cells was positively correlated with SLEDAI (p-value ≤ 0.05). A similar positive correlation pattern was observed for IKZF2 on non-classical monocytes and myeloid DCs, as well as IKZF3 on B cells ($0.05 <$ p-value ≤ 0.10).

In summary, a comprehensive immunophenotyping of target molecules' protein expression to scrutinize the subtle differences across discrete immune subsets provides insights into the functions of target molecules in the pathogenesis of SLE.

6.1 Gating strategy for immune sub-populations

Based on the promising SLE GWAS signals, I aimed to set up an antibody panel for the analysis of differential expression of IKZF1, IKZF2, IKZF3 and OX40L in innate immune cells (monocytes, dendritic cells (DCs), and NK cells) and adaptive immune cells (T cells and B cells) from patients and healthy controls. The experimental design is showed in Chapter 2 Methods, Table 2-5.

Gating of sample starts with the FSC-A and SSC-A to exclude the cell debris. To exclude dead cells, a 'viable gate' was set, followed by gating on SSC-A and SSC-W for single cells, which were PBMC (peripheral blood mononuclear cells). For panel 1 (Figure 6-1), I drew gates for the CD3+ population from PBMC and then discriminated the T cells into CD4 and CD8 subsets. They were then gated into Effector, Effector Memory, Central Memory and Naïve subpopulations by staining with CCR7 and CD45RA. Specifically, in CD4 subset, CCR6 and CXCR3 were used for gating the Th17, Th1, Th1Th17, and Th2 subpopulations. All the T cell subsets were then gated as the activated subpopulations with CD38+ and HLA-DR+ markers.

For regulatory T cells of panel 2 (Figure 6-2), I drew gates based on the CD3+ and CD4+ markers to get the CD4+ T cells subset. They were then gated on CCR4+ CD25+ and CCR6- to discriminate the regulatory T cell (Tregs) subsets. All the Tregs were then gated to separate Activated Treg, Memory Treg, and Naïve Treg subpopulations with CD45RO and HLA-DR markers.

For the B cells panel (Figure 6-3), gating on the CD3- and CD19+ was the B cell population. This was then discriminated into Non-switched Memory B cell, Double Negative B cell (DN B cell) by IgD and CD27 markers. The IgD-CD27+ population was then gated on the CD20 and CD38 to discriminate the plasmablasts and Switched Memory B cell. Another subset from PBMC was the CD3- CD20-, which was then separated as plasma cells by sequentially gating on the following markers: CD19-/CD138+/CD27high/CD38 high/IgD-.

For panel 4 (Figure 6-4), gating on the CD3+ and CD56+ from PBMC was the NKT cells. The gate on CD3- and CD19- was discriminated into CD14+ monocytes, which were further discriminated into classical and non-classical subpopulations by the CD16 marker. The CD14-

subset was gated to discriminate the NK cells by CD56+ and DCs by HLA-DR+. The DCs were then divided into Plasmacytoid DCs (CD123+) and Myeloid DCs (CD11c+).

In total, I have identified 51 cell subpopulations, four of which, i.e. Activated Effector CD4+ cell, Activated Effector CD8+ cell, Activated Central Memory CD8+ cell, and plasma cell, were removed due to their limited cells sample size (average cell counts of all samples < 100). Hence, 47 cell populations were used for the analysis of protein expression of target molecules. The markers for each immune cell type are summarized in Table 6-1. The expression levels of the target molecules are quantified by the median fluorescence intensity (MFI) of the corresponding antibody fluorophore conjugates, i.e., PE MFI for OX40L and FITC MFI for IKZF1, IKZF2, and IKZF3, as shown in Figure 6-5.

Table 6- 1. Surface markers for gating various cell populations.

Panel	Population Name	Corresponding Markers
Panel 1. T-cell	Total CD3+ T cells	CD3+
	Total CD4+ T cells	CD3+/CD4+/CD8-
	Activated CD4+ T cells	CD3+/CD4+/CD8-/CD38+/HLA-DR+
	Central Memory CD4+ T cells	CD3+/CD4+/CD8-/CCR7-CD45RA-
	Activated Central Memory CD4+ C	CD3+/CD4+/CD8-/CCR7-CD45RA-/CD38+/HLA-DR+
	Effector CD4+ T cells	CD3+/CD4+/CD8-/CCR7-/CD45RA+
	Effector Memory CD4+ T cells	CD3+/CD4+/CD8-/CCR7-/CD45RA-
	Activated Effector Memory Cd4+ T	CD3+/CD4+/CD8-/CCR7-/CD45RA-/CD38+/HLA-DR+
	Naive CD4+ T cells	CD3+/CD4+/CD8-/CCR7+/CD45RA+
	Th1	CD3+/CD4+/CD8-/CXCR3+/CCR6-
	Activated Th1	CD3+/CD4+/CD8-/CXCR3+/CCR6-/CD38+/HLA-DR+
	Th17	CD3+/CD4+/CD8-/CXCR3-/CCR6+
	Activated Th17	CD3+/CD4+/CD8-/CXCR3-/CCR6+/CD38+/HLA-DR+
	Th1Th17	CD3+/CD4+/CD8-/CXCR3+/CCR6+
	Activated Th1Th17	CD3+/CD4+/CD8-/CXCR3+/CCR6+/CD38+/HLA-DR+
	Th2	CD3+/CD4+/CD8-/CXCR3-/CCR6+
	Activated Th2	CD3+/CD4+/CD8-/CXCR3-/CCR6+/CD38+/HLA-DR+
	Total CD8+ T cells	CD3+/CD4-/CD8+
	Activated CD8+ T cells	CD3+/CD4-/CD8+/CD38+/HLA-DR+
	Central Memory CD8+ T cells	CD3+/CD4-/CD8+/CCR7-CD45RA-
	Effector CD8+ T cells	CD3+/CD4-/CD8+/CCR7-/CD45RA+
	Effector Memory CD8+ T cells	CD3+/CD4-/CD8+/CCR7-/CD45RA-
	Activated Effector Memory CD8+ T	CD3+/CD4-/CD8+/CCR7-/CD45RA-/CD38+/HLA-DR+
	Naive CD8+ T cells	CD3+/CD4-/CD8+/CCR7+/CD45RA+
	DP T cells	CD3+/CD4-/CD8-
Panel2. T-regulatory	Total T-regulatory	CD3+/CD4+/CD8-/CD25high/CCR4high/CD127low
	Activated T-regulatory	CD3+/CD4+/CD8-/CD25high/CCR4high/CD127low/HLA-DR+
	Memory T-regulatory	CD3+/CD4+/CD8-/CD25high/CCR4high/CD127low/CD45RO+
	Naïve T-regulatory	CD3+/CD4+/CD8-/CD25high/CCR4high/CD127low/CD45RO-
Panel3. B-cell	B cells	CD3-/CD19+
	DN B cells	CD3-/CD19-/IgD-/CD27-
	Naive B cells	CD3-/CD19+/IgD+/CD27-/CD24low/CD38low
	Transitional B cells	CD3-/CD19+/IgD+/CD27-/CD24high/CD38high
	Plasmablasts	CD3-/CD19+/IgD-/CD27+/CD20-/CD38+
	Switched Memory B cells	CD3-/CD19+/IgD-/CD27+/CD20+/CD38-
	Non-Switched Memory B cells	CD3-/CD19+/IgD+/CD27+
	Plasma cells	CD3-/CD19-/IgD-/CD20-/CD138+/CD27hi/CD38hi
Panel4. DC/Mono/NK	DCs	CD3-/CD19-/CD14-/HLA-DR+
	Myeloid DCs	CD3-/CD19-/CD14-/HLA-DR+/CD11c+/CD123-
	Plasmacytoid DCs	CD3-/CD19-/CD14-/HLA-DR+/CD11c-/CD123+
	NK cells	CD3-/CD19-/CD14-/CD56+
	CD16+ NK cells	CD3-/CD19-/CD14-/CD56+/CD16+
	CD16- NK cells	CD3-/CD19-/CD14-/CD56+/CD16-
	Monocytes	CD3-/CD19-/CD14+
	Classical Monocytes	CD3-/CD19-/CD14+/CD16-
	Non-classical Monocytes	CD3-/CD19-/CD14+/CD16+
	NKT cells	CD3+/CD56+

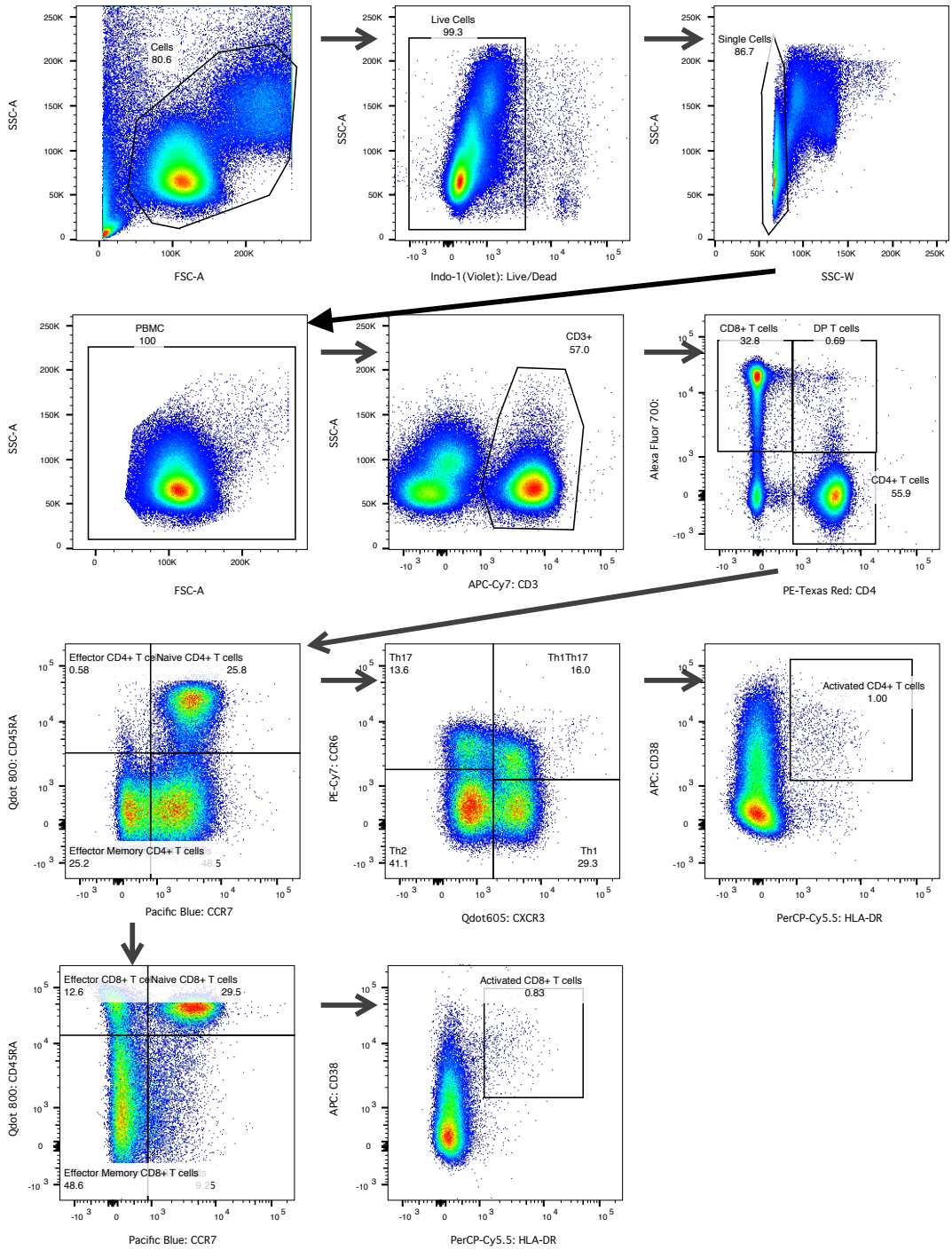


Figure 6- 1. Gating strategy of Panel 1 (T cell subsets).

Gating of the PBMC began with the FSC-A and SSC-A to exclude the cell debris. To exclude dead cells, a 'viable gate' was set, followed by gating on SSC-A and SSC-W for single cells. Gating the CD3+ (APC/Cy7) population from PBMC and the discrimination of the T cells into CD4 (PE-Texas Red) and CD8 (Alexa Fluor 700) subsets. They were then gated to Effector, Effector Memory, Central Memory and Naïve subpopulations by CCR7 (Pacific blue) and CD45RA (BV785). Specifically, in CD4 subset, CCR6 (PE/Cy7) and CXCR3 (BV605) were used for gating the Th17, Th1, Th1Th17, and Th2 subpopulations. All the T cell subsets were then gated as the activated subpopulations with CD38+ (APC) and HLA-DR+ (PerCP-Cy5.5) markers.

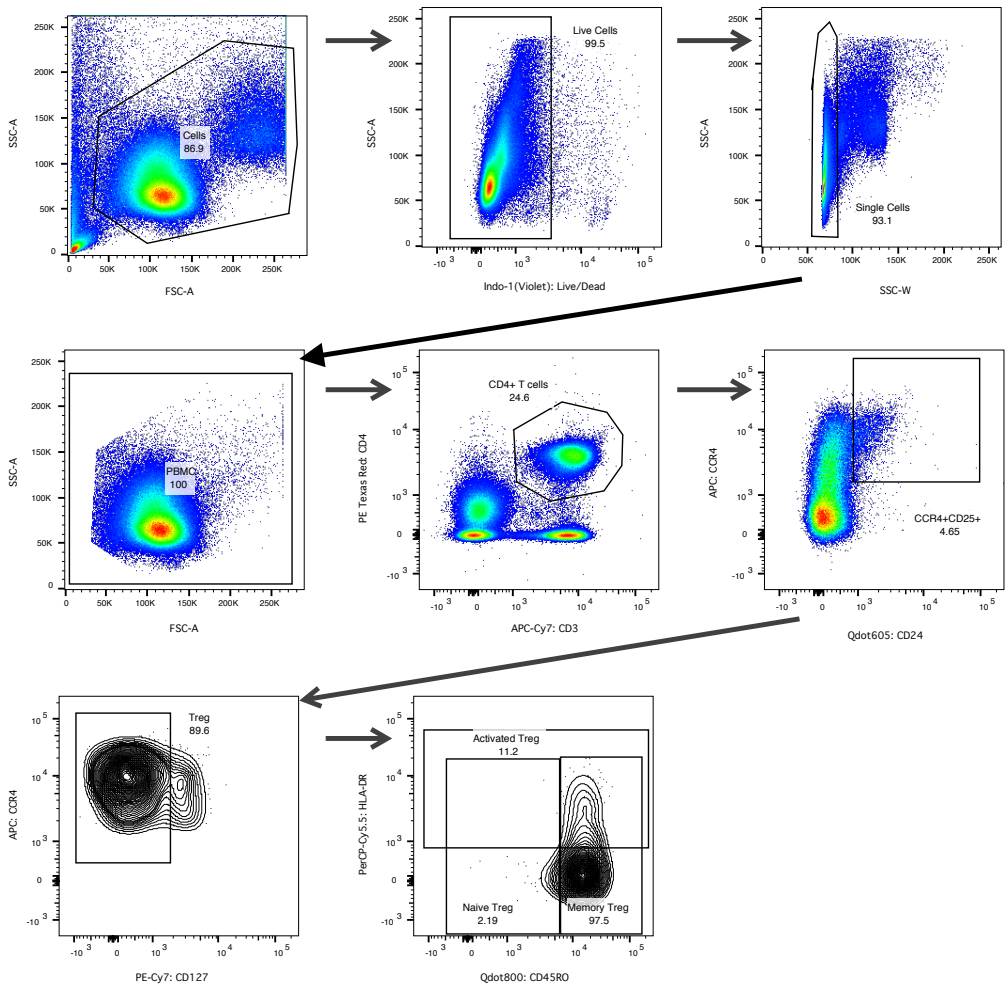


Figure 6- 2. Gating strategy of Panel 2 (Regulatory T cell subsets).

Gating of sample starts with the FSC-A and SSC-A to exclude the cell debris. Gating on the CD3+ (APC/Cy7) and CD4+ (PE-Texas Red) was the CD4+ T cells subset. They were then gating on CCR4+ (APC), CD25+ (Qdot 605) and CD127 low (PE-Cy7) to discriminate the Treg subsets. All the Tregs were then gated to Activate Treg, Memory Treg, and Naïve Treg subpopulations with CD45RO (Qdot 800) and HLA-DR (PerCP-Cy5.5) markers.

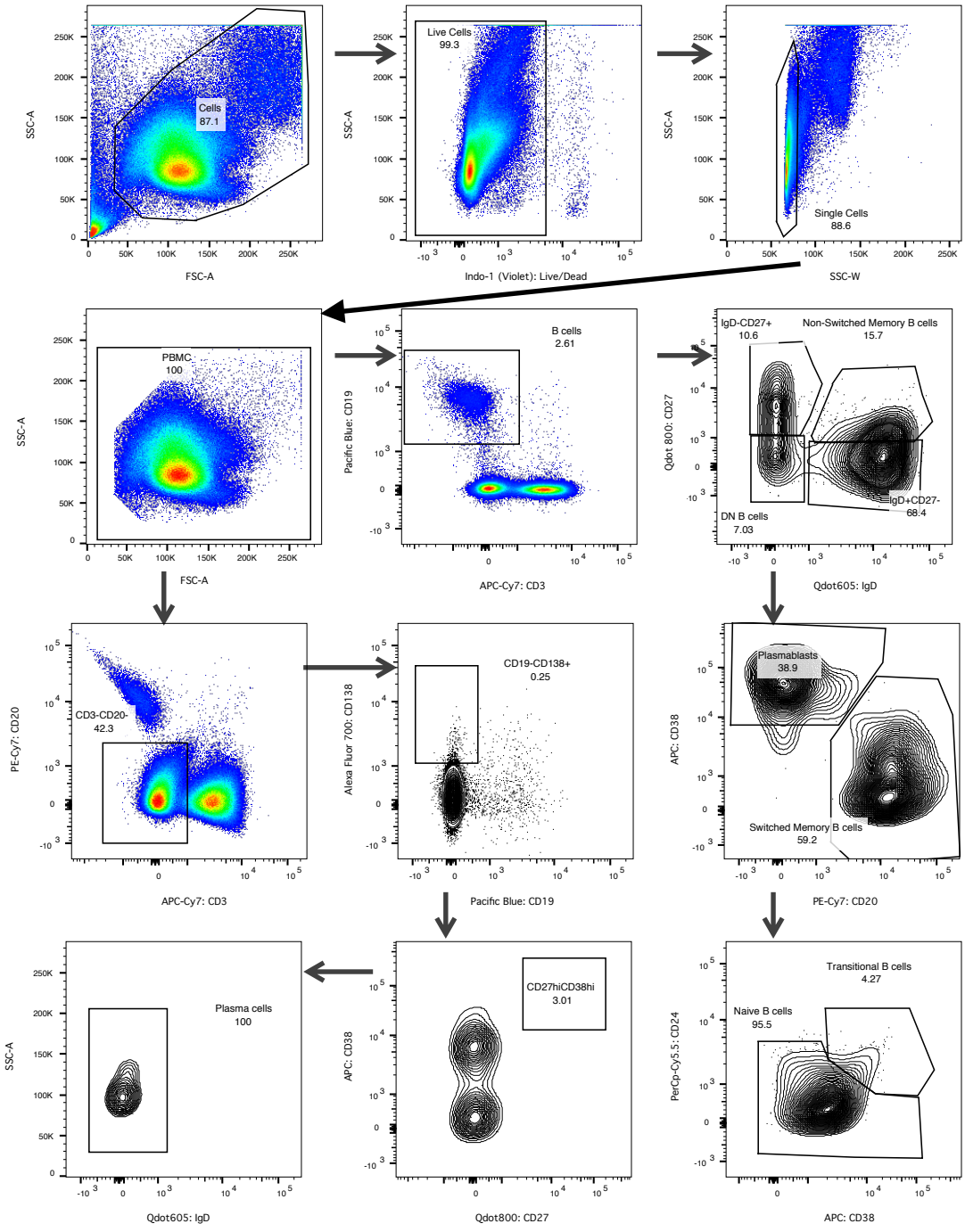


Figure 6- 3. Gating strategy of Panel 3 (B cell subsets).

Gating of sample starts with the FSC-A and SSC-A to exclude the cell debris. Gating on the CD3- (APC/Cy7) and CD19+ (Pacific blue) were B cell population. This was discriminated into Non-switched Memory B cell, DN B cell by IgD (Qdot 605) and CD27 (Qdot 800) markers. The 'IgD-CD27+' population was then gating on the CD20 (PE-Cy7) and CD38 (APC) to discriminate the plasmablasts and Switched Memory B cell. The 'IgD+CD27-' population was gating on CD38 (APC) and CD24 (PerCP-Cy5.5) to discriminate the Naïve B cells and the Transitional B cells. Another subset from PBMC was the CD3- (APC-Cy7) CD20- (PE-Cy7), which was then discriminated into plasma cells by CD19- (Pacific blue) CD138+ (Alexa Fluor 700), CD27high (Qdot 800), CD38 high (APC) and IgD- (Qdot 605) markers.

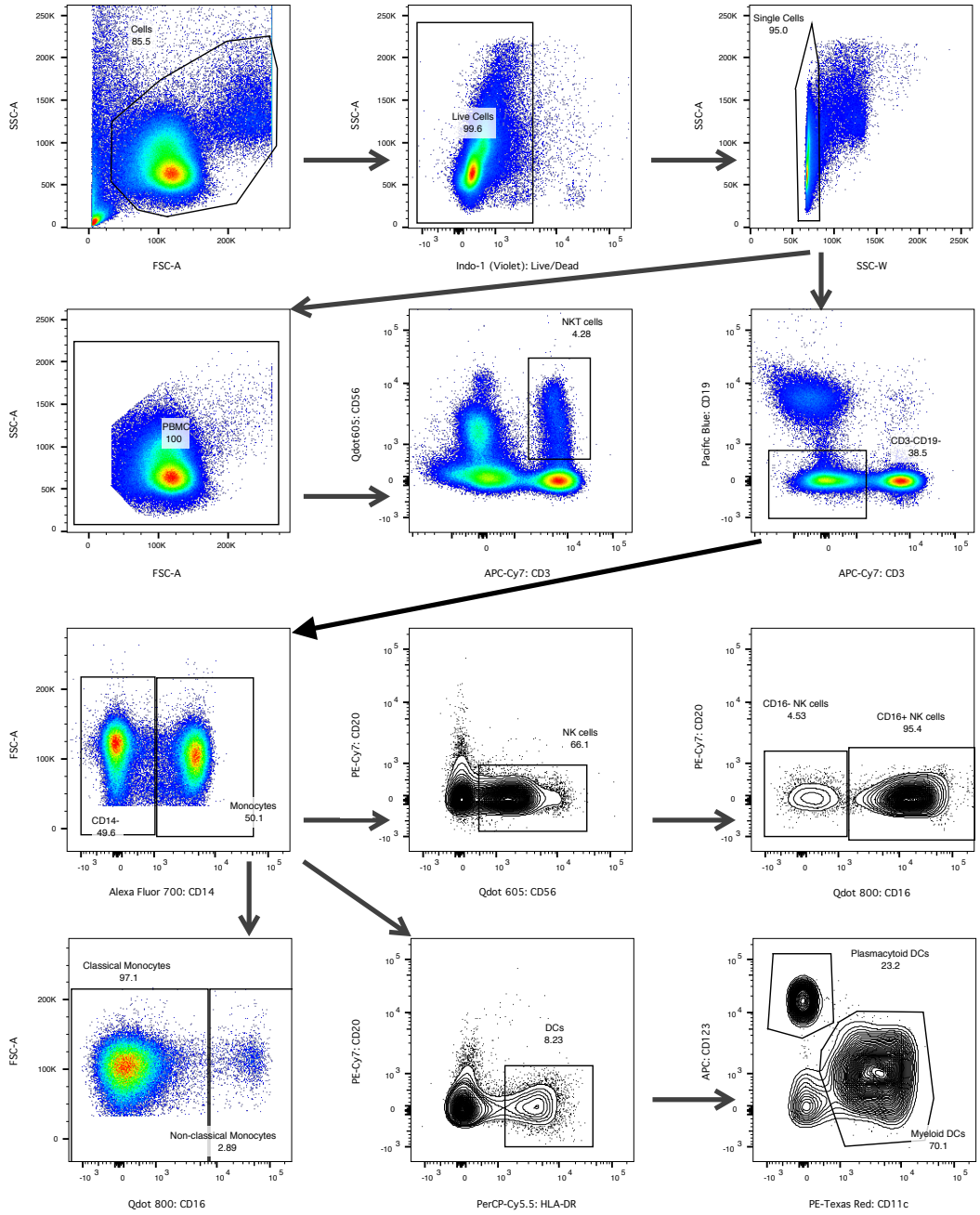


Figure 6- 4. Gating strategy of Panel 4 (NK, DC, and Monocytes).

Gating of sample starts with the FSC-A and SSC-A to exclude the cell debris. Gating on the CD3+ (APC/Cy7) and CD56+ (Qdot 605) from PBMC was the NKT cell. The gate on CD3- (APC/Cy7) and CD19- (Pacific blue) was discriminated into CD14+ (Alexa Fluor 700) monocytes, which were further discriminated into classical and non-classical subpopulations by CD16 (Qdot 800) marker. The CD14- (Alexa 700) subset was gated to discriminate the NK cells (CD56+ Qdot 605) and DCs (HLA-DR+ PerCp-Cy5.5). The DCs were then divided into plasmacytoid DCs (CD123+ APC) and Myeloid DCs (CD11c+ PE-Texas Red).

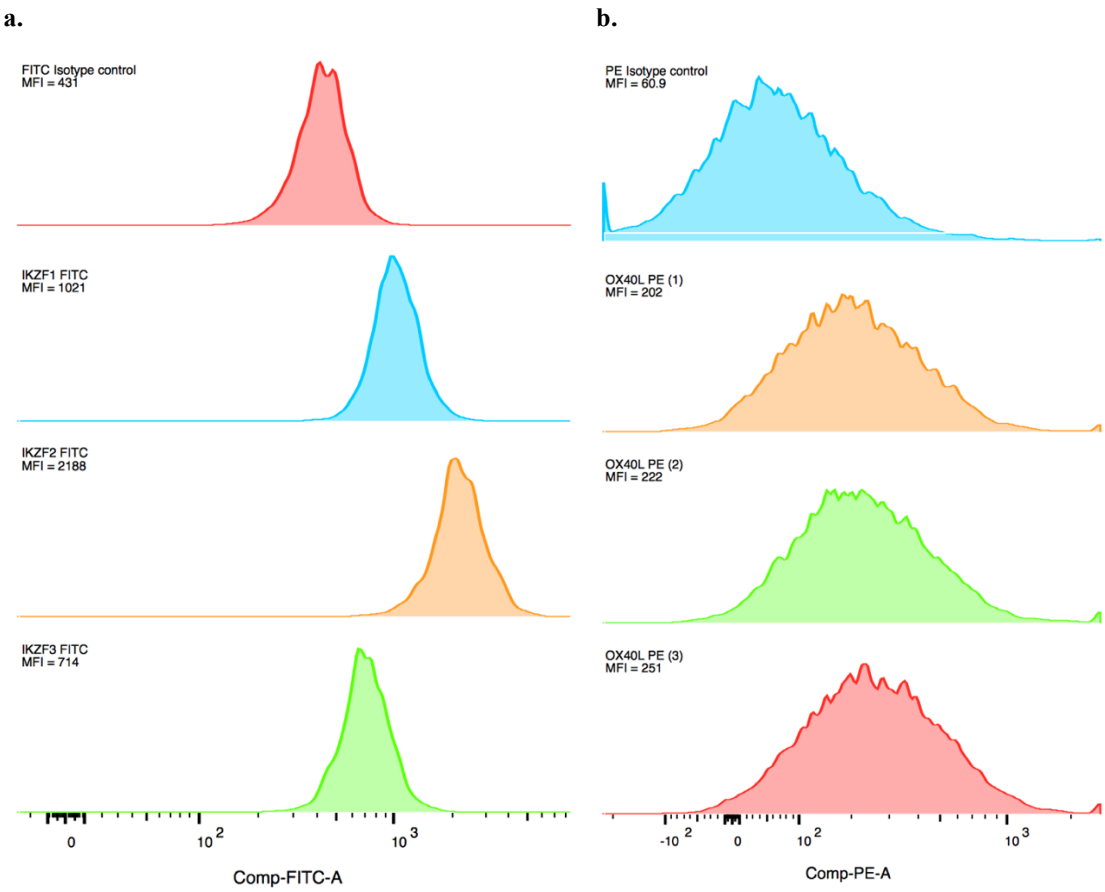


Figure 6- 5. **Quantification of protein expression of target molecules.**
Left panel - IKZF (FITC) expression signals on Treg subset. From top to down, the histograms were isotype control (red), IKZF1 (blue), IKZF2 (yellow), and IKZF3 (green). The protein level of IKZF1/2/3 was calculated by the following formula: *MFI_IKZF – MFI_IsotypeControl*.
Right panel - OX40L (PE) expression on B cells subset. The top right histogram was isotype control PE (blue) and the remaining three histograms were for OX40L PE. The protein level of OX40L was calculated by the following formula: *(MFI_OX40L(1) + MFI_OX40L(2))/2 – MFI_IsotypeControl*. Histograms with the same colour on the left and right panel indicate the corresponding FITC (IKZF1/2/3 or isotype control) and PE (OX40L or isotype control) were added to the same sample.

6.2 The cell counts of immune subsets in PBMC

PBMC is a bulk tissue comprising many cell types, including B and T lymphocytes, monocytes, dendritic cells, and natural killer cells. I analysed these composition data from flow cytometry using R and GraphPad Prism 7 to test if there is any difference between the immune cells composition in PBMC between patients and controls. In total, 51 immune subsets were identified, while four of them, including Activated Effector CD4+ T cells, Activated Central Memory CD8+ T cells, Activated Effector CD8+ T cells, and plasma cells, were removed due to their relative small sample size, leaving 47 immune subsets for further analyses. Figure 6-6 and Table 6-2 shows the summary statistics for the cell counts of all the 47 immune subsets and the absolute number of PBMC. When examining the cell counts of these cell subsets between 22 SLE patients and 20 healthy controls, I observed that the absolute number of PBMCs was decreased (P= 0.0196) in SLE patients (Mean = 248206) when comparing to healthy controls (Mean = 270539). Moreover, I found that immune subsets who had a significant decreased in the absolute cell counts in SLE patients were majorly innate immune cells, including NK cells (mean difference (95% CI): -14951 (-23473, -6430); P = 0.00108), NKT cells (-9494 (-14223, -4765); P = 0.000348), and Plasmacytoid DCs (-724 (-1141, -308); P = 0.00134), all with a FDR < 0.05. While the cell counts of adaptive immune subsets were more likely to be elevated in SLE patients when comparing to controls, such as Th2 (16337 (188, 32486); P = 0.0476), Activated Th2 (186 (86, 285); P= 0.000779), Activated Th1 (202 (27, 377); P = 0.0256), Activated CD4+ T cells (508 (165, 850); P = 0.00499), Activated Central Memory CD4+ T cells (226 (42, 411); P = 0.0185), as well as DN B cells (963 (150, 1776); P = 0.0224). Interestingly, the absolute number of Th1Th17 (-6865 (-10859, -2870); P = 0.00125) and Effector Memory CD8+ T cells (-5958 (-11887, -29); P = 0.0489) were decreased in SLE patients. When examining the correlation between the cell counts of immune subsets and disease activity as measured by SLEDAI, I found that the absolute numbers of both monocytes (P = 0.055) and classical monocytes (P = 0.046) were positively correlated with SLEDAI [157] (Figure 6-7. a & b), while both naïve B cells (P = 0.038) and non-switched memory B cells (P = 0.0036) were negatively correlated with SLEDAI (Figure 6-7. c & d), indicating the innate and adaptive immune system may have distinct functions in the activeness of SLE.

Table 6- 2. The absolute number of immune subsets.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	PBMC	248206	270539	-22333 (-40774, -3893)	0.0196	0.105
PANEL1	CD3+ T cells	145362	161786	-16424 (-39644, 6795)	0.158	0.323
PANEL1	CD4+ T cells	102197	102497	-300 (-22694, 22094)	0.978	0.978
PANEL1	Central Memory CD4+ T cells	41282	47837	-6555 (-19192, 6082)	0.3	0.427
PANEL1	Effector CD4+ T cells	1454	1823	-369 (-1745, 1007)	0.586	0.689
PANEL1	Effector Memory CD4+ T cells	13361	16373	-3013 (-8650, 2625)	0.286	0.42
PANEL1	Naive CD4+ T cells	45749	36463	9286 (-6752, 25324)	0.246	0.385
PANEL1	CD8+ T cells	34710	41578	-6868 (-21376, 7640)	0.344	0.476
PANEL1	Central Memory CD8+ T cells	2983	3512	-529 (-2332, 1273)	0.556	0.67
PANEL1	Effector CD8+ T cells	7913	14481	-6568 (-14265, 1129)	0.092	0.24
PANEL1	Effector Memory CD8+ T cells	6597	12555	-5958 (-11887, -29)	0.0489	0.192
PANEL1	Naive CD8+ T cells	17210	11031	6179 (-309, 12668)	0.0611	0.205
PANEL1	DP T cells	1540	1219	321 (-419, 1061)	0.383	0.514
PANEL1	Th1	20297	27020	-6723 (-15722, 2276)	0.139	0.323
PANEL1	Th17	11045	14095	-3050 (-7250, 1150)	0.15	0.323
PANEL1	Th1Th17	5667	12531	-6865 (-10859, -2870)	0.00125	0.0105
PANEL1	Th2	65189	48851	16337 (188, 32486)	0.0476	0.192
PANEL1	Activated CD4+ T cells	1159	651	508 (165, 850)	0.00499	0.0335
PANEL1	Activated Central Memory CD4+ T cells	456	230	226 (42, 411)	0.0185	0.109
PANEL1	Activated Effector Memory CD4+ T cells	250	112	138 (-2, 279)	0.0537	0.194
PANEL1	Activated Th1	407	205	202 (27, 377)	0.0256	0.12
PANEL1	Activated Th17	152	106	46 (-29, 122)	0.218	0.381
PANEL1	Activated Th1Th17	120	104	16 (-31, 64)	0.493	0.626
PANEL1	Activated Th2	282	96	186 (86, 285)	0.000779	0.0105
PANEL1	Activated CD8+ T cells	677	322	355 (-145, 855)	0.155	0.323
PANEL1	Activated Central Memory CD8+ T cells	90	47	43 (-4, 90)	0.0705	0.207
PANEL1	Activated Effector Memory CD8+ T cells	257	117	140 (-115, 395)	0.266	0.403
PANEL2	Treg	2427	2051	376 (-249, 1000)	0.228	0.381
PANEL2	Memory Treg	2283	1929	355 (-243, 952)	0.235	0.381
PANEL2	Naive Treg	139	119	20 (-37, 77)	0.481	0.626
PANEL2	Activated Treg	362	251	111 (-72, 294)	0.221	0.381
PANEL3	B cells	15736	10183	5553 (-609, 11715)	0.075	0.207
PANEL3	DN B cells	1745	782	963 (150, 1776)	0.0224	0.117
PANEL3	Naive B cells	7853	4764	3089 (-1389, 7566)	0.166	0.325
PANEL3	Transitional B cells	828	371	457 (-43, 957)	0.0711	0.207
PANEL3	Plasmablasts	1038	442	596 (-191, 1383)	0.13	0.322
PANEL3	Switched Memory B cells	1693	1702	-9 (-594, 576)	0.975	0.978
PANEL3	Non-Switched Memory B cells	2508	2039	469 (-1360, 2299)	0.604	0.692
PANEL4	DCs	6656	6930	-274 (-2865, 2317)	0.832	0.899
PANEL4	Myeloid DCs	5697	5412	285 (-2122, 2693)	0.811	0.899
PANEL4	Plasmacytoid DCs	533	1258	-724 (-1141, -308)	0.00134	0.0105
PANEL4	NK cells	19615	34566	-14951 (-23473, -6430)	0.00108	0.0105
PANEL4	CD16+ NK cells	17702	32469	-14766 (-23230, -6303)	0.00116	0.0105
PANEL4	CD16- NK cells	1898	2081	-183 (-744, 378)	0.513	0.635
PANEL4	Monocytes	42018	40662	1355 (-12410, 15120)	0.842	0.899
PANEL4	Classical Monocytes	37760	37999	-239 (-12392, 11913)	0.968	0.978
PANEL4	Non-classical Monocytes	4219	2634	1585 (-874, 4045)	0.197	0.37
PANEL4	NKT cells	3930	13424	-9494 (-14223, -4765)	0.000348	0.0105

Notes: colours in the table: 1) red indicates a significant difference of cell proportions between SLE and controls with a p-value ≤ 0.05 & q-value ≤ 0.05; 2) purple indicates a p-value ≤ 0.05 & q-value > 0.05.

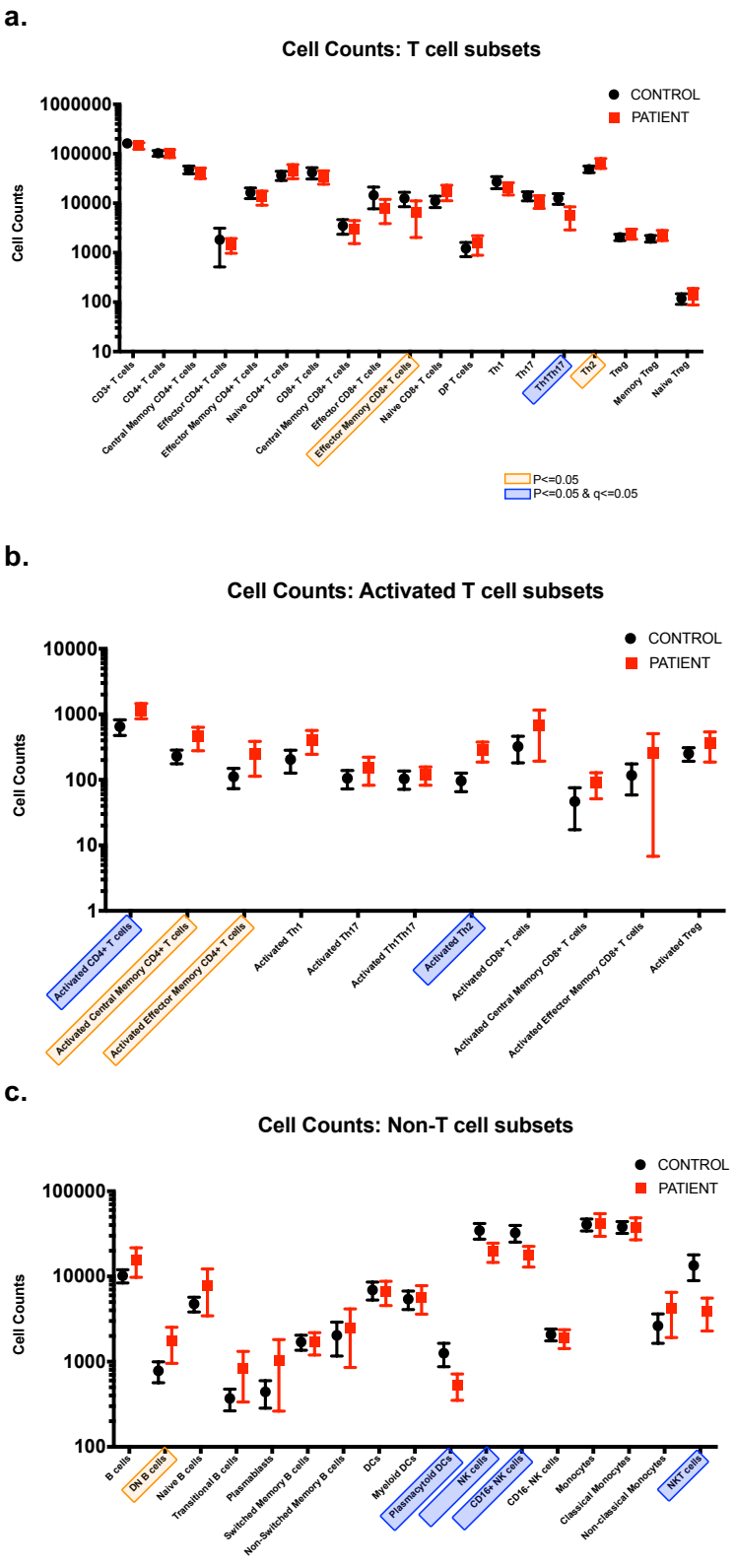


Figure 6- 6. **The absolute number of immune subsets in patients and controls.**

Top panel: T cell subsets; Middle panel: Active T cell subsets; Bottom panel: monocytes, NK cells and DCs. Cell types highlighted with light yellow boxes have passed the multiple testing adjustment with a q value $\leqslant 0.05$. Cell types highlighted with blue boxes have a p value $\leqslant 0.05$ but failed to passed the multiple testing adjustment. Y-axis is presented in a log10 scale. The cell counts are showed as mean and 95%CI.

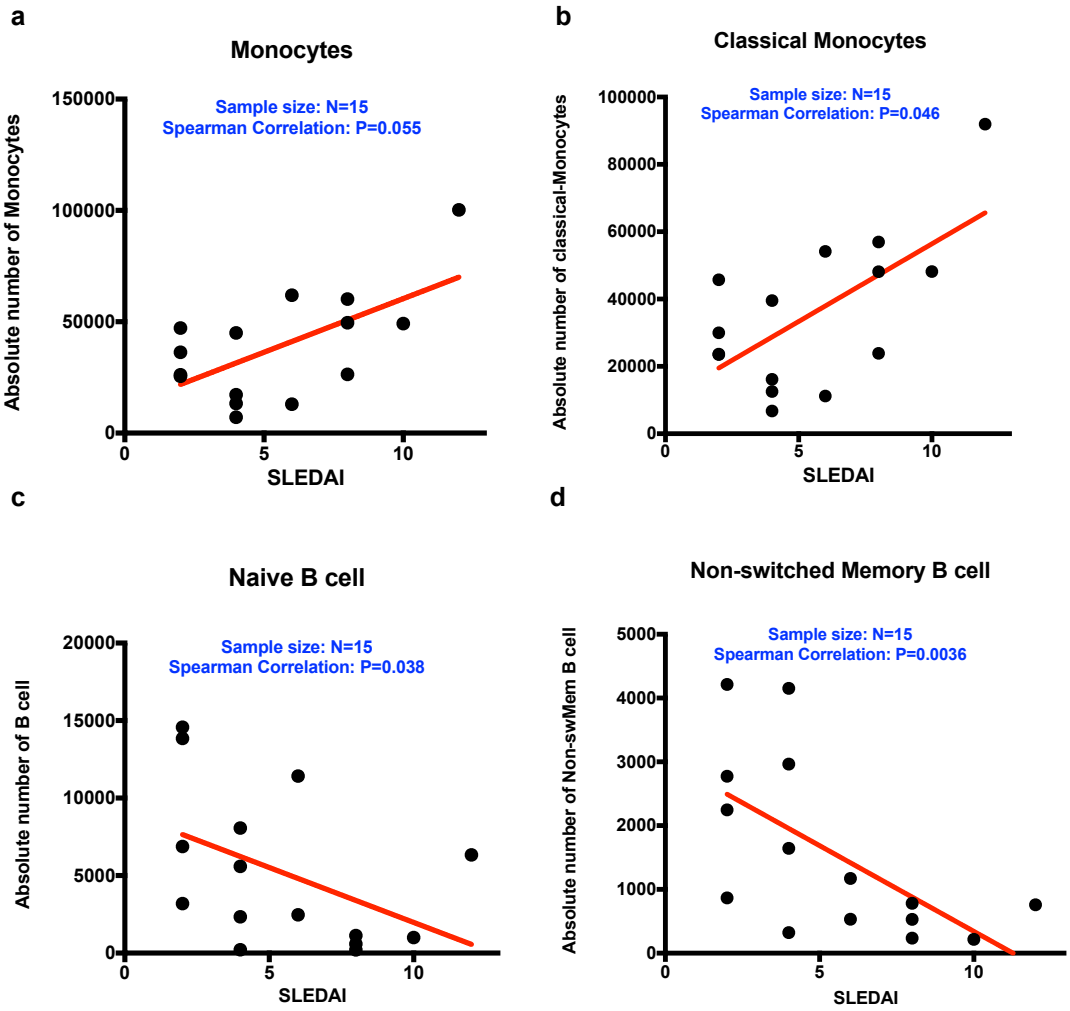


Figure 6- 7. **Correlation between the absolute number of immune subsets and SLEDAI.** Top panel: Positive correlation of SLEDAI with absolute number of Monocytes (a) and Classical monocytes (b). Bottom panel: Negative correlation of SLEDAI with absolute number of Naïve B cells (c) and Non-switched Memory B cells (d).

6.3 The cell proportions of immune subsets in PBMC

Figure 6-8 and Table 6-4 shows the summary statistics for the proportions (percentage in PBMC) of all the 47 immune subsets. When examining the relative frequencies of cell subsets between 22 SLE patients and 20 healthy controls, I observed that Th2 cells were elevated (7.98 (2.2, 13.76), P = 0.00853), while Th1Th17 cells were decreased (-2.39 (-3.81, -0.97), P = 0.00155) in SLE patients compared to healthy controls. When looking at the activated T cell panel, the frequencies of activated T cell subpopulations were higher in SLE than healthy controls, including activated central memory CD4+ T cell (0.1 (0.03, 0.18), P = 0.0104), activated effector memory CD4+ T cells (0.06 (0.01, 0.12), P = 0.0296), activated Th1 cells (0.09 (0.02, 0.16), P = 0.0138) and activated Th2 cells 0.08 (0.04, 0.13), P = 0.000812). Notably, the proportion of B cells (2.65 (0.25, 5.05), P = 0.032) and transitional B cells (0.21 (0, 0.41), P = 0.0461) were higher in SLE than in healthy control, while the proportion of NK cells (-4.03 (-7.52, -0.54), P = 0.0247) and Plasmacytoid DCs (-0.25 (-0.39, -0.1), P = 0.00188) was higher in controls than in patients When examining the correlation between the frequencies of immune subsets and disease activity as measured by SLEDAI [157], I found that NK cell frequencies and B cell frequencies were both negatively correlated with SLEDAI, suggesting that patients with more active SLE tended to have more severe leukopenia, showed in Figure 6-9.

Table 6- 3. Proportion of immune cells in PBMC.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	CD3 T cells	57.94	59.85	-1.91 (-9.64, 5.83)	0.618	0.687
PANEL1	CD4+ T cells	40.66	37.86	2.81 (-5.07, 10.68)	0.475	0.609
PANEL1	**Activated CD4+ T cells**	**0.48**	**0.24**	**0.24 (0.09, 0.39)**	**0.00236**	**0.0236**
PANEL1	Central Memory CD4+ T cells	16.61	17.68	-1.07 (-5.78, 3.64)	0.648	0.704
PANEL1	**Activated Central Memory CD4+ T cells**	**0.19**	**0.09**	**0.1 (0.03, 0.18)**	**0.0104**	**0.0743**
PANEL1	Effector CD4+ T cells	0.6	0.69	-0.09 (-0.63, 0.45)	0.741	0.756
PANEL1	*Activated Effector CD4+ T cells*	*0.03*	*0.01*	*0.03 (-0.02, 0.07)*	*0.231*	*0.374*
PANEL1	Effector Memory CD4+ T cells	5.41	6.09	-0.68 (-2.82, 1.45)	0.521	0.61
PANEL1	**Activated Effector Memory CD4+ T cells**	**0.1**	**0.04**	**0.06 (0.01, 0.12)**	**0.0296**	**0.114**
PANEL1	Naive CD4+ T cells	17.91	13.4	4.51 (-1.27, 10.29)	0.121	0.238
PANEL1	Th1	8.04	10.07	-2.03 (-5.43, 1.36)	0.232	0.374
PANEL1	**Activated Th1**	**0.17**	**0.08**	**0.09 (0.02, 0.16)**	**0.0138**	**0.0767**
PANEL1	Th17	4.43	5.18	-0.75 (-2.29, 0.8)	0.334	0.464
PANEL1	Activated Th17	0.07	0.04	0.03 (-0.02, 0.08)	0.196	0.35
PANEL1	**Th1Th17**	**2.21**	**4.6**	**-2.39 (-3.81, -0.97)**	**0.00155**	**0.0235**
PANEL1	Activated Th1Th17	0.05	0.04	0.01 (-0.01, 0.03)	0.244	0.381
PANEL1	**Th2**	**25.98**	**18**	**7.98 (2.2, 13.76)**	**0.00853**	**0.0711**
PANEL1	**Activated Th2**	**0.12**	**0.04**	**0.08 (0.04, 0.13)**	**0.000812**	**0.0203**
PANEL1	CD8+ T cells	13.82	15.45	-1.64 (-6.92, 3.65)	0.535	0.61
PANEL1	Activated CD8+ T cells	0.28	0.12	0.16 (-0.03, 0.35)	0.103	0.238
PANEL1	Central Memory CD8+ T cells	1.19	1.32	-0.13 (-0.82, 0.56)	0.706	0.735
PANEL1	*Activated Central Memory CD8+ T cells*	*0.04*	*0.02*	*0.02 (0, 0.04)*	*0.0447*	*0.136*
PANEL1	Effector CD8+ T cells	3.13	5.37	-2.24 (-5.06, 0.58)	0.116	0.238
PANEL1	*Activated Effector CD8+ T cells*	*0.05*	*0.03*	*0.02 (-0.02, 0.07)*	*0.315*	*0.45*
PANEL1	Effector Memory CD8+ T cells	2.58	4.67	-2.09 (-4.29, 0.12)	0.063	0.175
PANEL1	Activated Effector Memory CD8+ T cells	0.1	0.04	0.06 (-0.04, 0.16)	0.212	0.366
PANEL1	**Naive CD8+ T cells**	**6.9**	**4.09**	**2.81 (0.39, 5.24)**	**0.0246**	**0.103**
PANEL1	DP T cells	0.66	0.45	0.21 (-0.16, 0.58)	0.259	0.392
PANEL2	Treg	0.97	0.74	0.23 (-0.03, 0.49)	0.0818	0.215
PANEL2	Activated Treg	0.14	0.09	0.05 (-0.02, 0.12)	0.124	0.238
PANEL2	Memory Treg	0.92	0.7	0.22 (-0.03, 0.47)	0.0867	0.217
PANEL2	Naive Treg	0.05	0.04	0.01 (-0.01, 0.03)	0.267	0.393
PANEL3	**B cells**	**6.36**	**3.72**	**2.65 (0.25, 5.05)**	**0.032**	**0.114**
PANEL3	**DN B cells**	**0.71**	**0.29**	**0.42 (0.1, 0.74)**	**0.0123**	**0.0767**
PANEL3	Naive B cells	3.11	1.74	1.37 (-0.38, 3.11)	0.118	0.238
PANEL3	**Transitional B cells**	**0.34**	**0.13**	**0.21 (0, 0.41)**	**0.0461**	**0.136**
PANEL3	Plasmablasts	0.47	0.16	0.31 (-0.08, 0.69)	0.11	0.238
PANEL3	Switched Memory B cells	0.68	0.62	0.06 (-0.14, 0.27)	0.537	0.61
PANEL3	Non Switched Memory B cells	1.02	0.74	0.28 (-0.41, 0.97)	0.415	0.546
PANEL3	*Plasma cells*	*0.02*	*0*	*0.02 (0, 0.04)*	*0.0361*	*0.12*
PANEL4	DCs	2.62	2.5	0.12 (-0.83, 1.07)	0.795	0.795
PANEL4	Myeloid DCs	2.24	1.95	0.29 (-0.6, 1.18)	0.514	0.61
PANEL4	**Plasmacytoid DCs**	**0.21**	**0.45**	**-0.25 (-0.39, -0.1)**	**0.00188**	**0.0235**
PANEL4	**NK cells**	**8.36**	**12.39**	**-4.03 (-7.52, -0.54)**	**0.0247**	**0.103**
PANEL4	**CD16+ NK cells**	**7.54**	**11.63**	**-4.09 (-7.53, -0.66)**	**0.0209**	**0.103**
PANEL4	CD16- NK cells	0.81	0.75	0.06 (-0.23, 0.35)	0.664	0.706
PANEL4	Monocytes	17.18	14.7	2.49 (-3.35, 8.32)	0.389	0.526
PANEL4	Classical Monocytes	15.35	13.72	1.63 (-3.28, 6.55)	0.501	0.61
PANEL4	Non-classical Monocytes	1.81	0.97	0.85 (-0.35, 2.05)	0.158	0.293
PANEL4	**NKT cells**	**1.58**	**4.81**	**-3.22 (-4.93, -1.52)**	**0.000599**	**0.0203**

Notes: colours in the table: 1) red indicates a significant difference of cell proportions between SLE and controls with a p-value ≤ 0.05 & q-value ≤ 0.05; 2) purple indicates a p-value ≤ 0.05 & q-value > 0.05; 3) grey and italic indicates the cell subpopulation is excluded due to the limited number of cell counts.

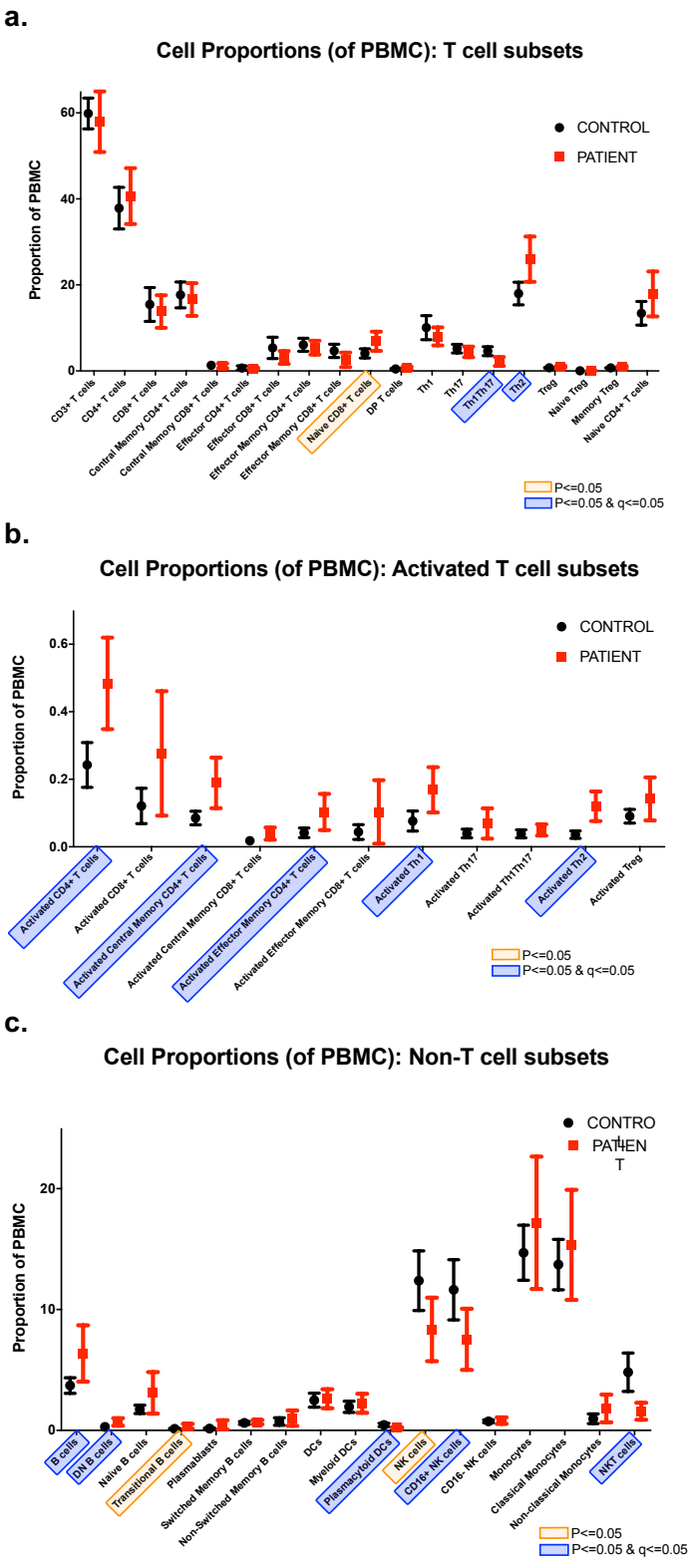


Figure 6- 8. **Immune cells composition of PBMC in patients and controls.**

Top panel: T cell subsets; Middle panel: Active T cell subsets; Bottom panel: monocytes, NK cells and DCs. Cell types highlighted with light yellow boxes have passed the multiple testing adjustment with a q value $\leqslant 0.05$. Cell types highlighted with blue boxes have a p value $\leqslant 0.05$ but failed to passed the multiple testing adjustment. The cell proportions are showed as mean and 95%CI.

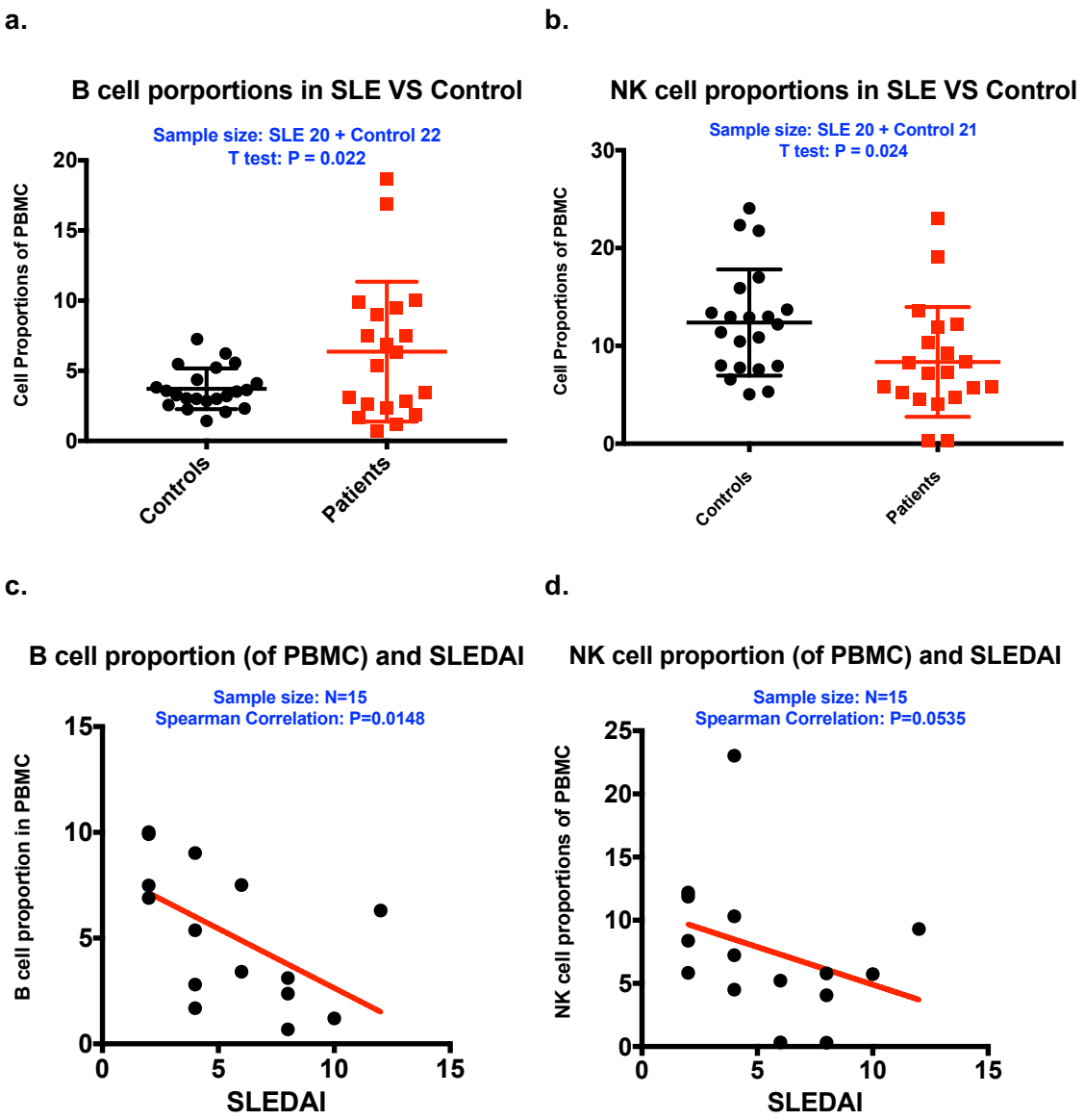


Figure 6- 9. **Correlation between the cell proportions of immune subsets and SLEDAI.** Left panel. B cell proportions in SLE patients and healthy controls (a) and the correlation of SLEDAI and B cell proportions (c). Right panel: NK cell proportions in SLE patients and healthy controls (b) and the correlation of SLEDAI and NK cell proportions (d).

6.4 OX40L protein expression

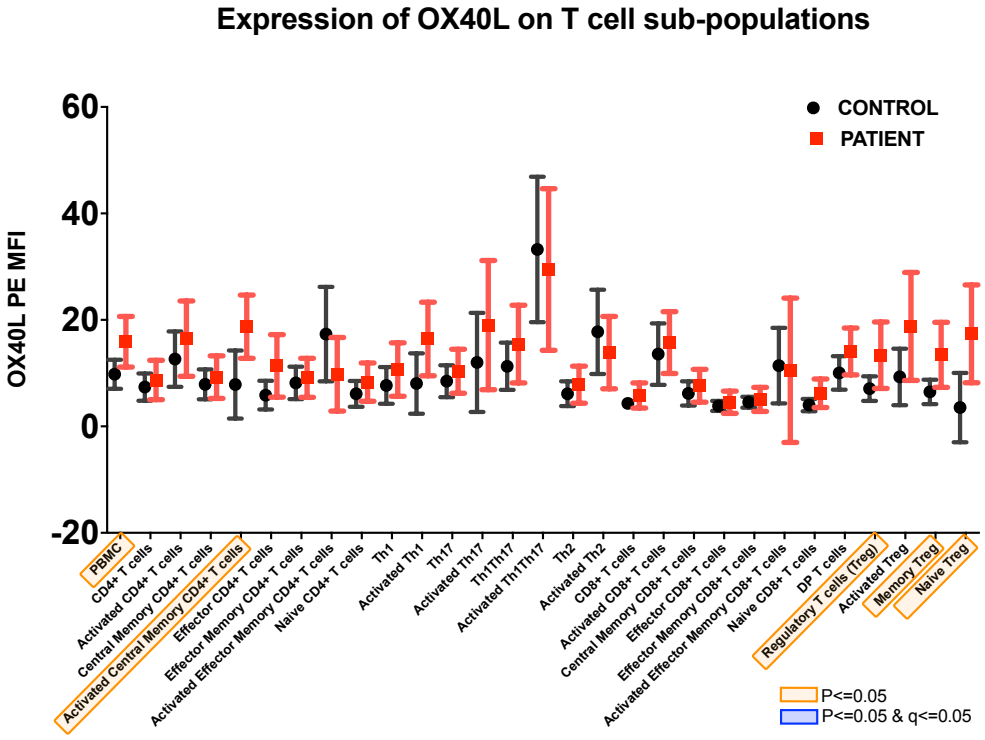
Figure 6-10 and Table 6-4 showed the average expression level of the OX40L in patients and controls among all cell types being analysed. The expression level of OX40L was relatively low in all T cell subsets, with a MFI (median fluorescence intensity) ranged from 5 to 20, while the MFI on the activated Th1Th17 was obviously higher than others, ranging from 30 to 40. When looking at the non-T cell panel showed in Figure 6-8, the expression of OX40L was much higher in patients than that in controls on B cells (48.5 (0.9, 96.2), P = 0.046) and Non-classical monocytes (32.9 (0.7, 65.1), P = 0.0455), indicating the protein level of OX40L was correlated with the disease status. Previous studies have showed that the level of OX40L is significantly increased under the activation of immune system [201]. Moreover, OX40L is found to exclusively express on APCs (Antigen presenting cells, APC) and especially on activated cells, consistent with the findings from this study [121].

Table 6- 4. Average level of OX40L on immune subpopulations in SLE and controls.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	**PBMC**	**15.9**	**9.8**	**6.1 (0.8, 11.4)**	**0.0264**	**0.283**
PANEL1	CD3 T cells	7.2	5.9	1.4 (-1.9, 4.6)	0.396	0.554
PANEL1	CD4+ T cells	8.7	7.4	1.3 (-3, 5.7)	0.542	0.658
PANEL1	Activated CD4+ T cells	16.5	12.6	3.9 (-4.7, 12.4)	0.365	0.532
PANEL1	Central Memory CD4+ T cells	9.3	7.9	1.4 (-3.4, 6.1)	0.567	0.672
PANEL1	**Activated Central Memory CD4+ T cells**	**18.7**	**7.9**	**10.9 (2.4, 19.3)**	**0.013**	**0.283**
PANEL1	Effector CD4+ T cells	11.4	5.9	5.5 (-0.8, 11.8)	0.0863	0.297
PANEL1	*Activated Effector CD4+ T cells*	*15.1*	*38.1*	*-23 (-52.8, 6.8)*	*0.126*	*0.326*
PANEL1	Effector Memory CD4+ T cells	9.1	8.2	1 (-3.7, 5.6)	0.676	0.734
PANEL1	Activated Effector Memory CD4+ T cells	9.8	17.3	-7.6 (-18.5, 3.4)	0.169	0.359
PANEL1	Naive CD4+ T cells	8.3	6.1	2.2 (-2, 6.4)	0.297	0.485
PANEL1	Th1	10.7	7.7	3 (-2.9, 8.9)	0.311	0.485
PANEL1	Activated Th1	16.4	8.1	8.4 (-0.3, 17)	0.058	0.283
PANEL1	Th17	10.4	8.5	1.9 (-3.1, 6.8)	0.452	0.576
PANEL1	Activated Th17	19	12	7 (-7.8, 21.8)	0.345	0.518
PANEL1	Th1Th17	15.5	11.3	4.2 (-4.1, 12.5)	0.314	0.485
PANEL1	Activated Th1Th17	29.5	33.2	-3.8 (-23.5, 16)	0.702	0.746
PANEL1	Th2	7.8	6.1	1.7 (-2.4, 5.8)	0.402	0.554
PANEL1	Activated Th2	13.9	17.8	-3.9 (-14, 6.2)	0.439	0.574
PANEL1	CD8+ T cells	5.8	4.3	1.5 (-1, 4)	0.238	0.434
PANEL1	Activated CD8+ T cells	15.7	13.6	2.2 (-5.8, 10.1)	0.582	0.675
PANEL1	Central Memory CD8+ T cells	7.6	6.2	1.5 (-2.3, 5.2)	0.432	0.574
PANEL1	*Activated Central Memory CD8+ T cells*	*23.6*	*7.9*	*15.7 (-1.4, 32.8)*	*0.071*	*0.283*
PANEL1	Effector CD8+ T cells	4.5	3.8	0.7 (-1.6, 2.9)	0.539	0.658
PANEL1	*Activated Effector CD8+ T cells*	*16.4*	*14*	*2.4 (-16.2, 20.9)*	*0.797*	*0.83*
PANEL1	Effector Memory CD8+ T cells	5.1	4.5	0.5 (-1.9, 3)	0.656	0.727
PANEL1	Activated Effector Memory CD8+ T cells	10.6	11.4	-0.9 (-15.8, 14.1)	0.907	0.907
PANEL1	Naive CD8+ T cells	6.2	4	2.2 (-0.7, 5.1)	0.126	0.326
PANEL1	DP T cells	14.1	10	4 (-1.2, 9.3)	0.128	0.326
PANEL2	Treg	13.4	7.1	6.3 (-0.3, 12.9)	0.0601	0.283
PANEL2	Activated Treg	18.8	9.3	9.5 (-1.7, 20.7)	0.0933	0.297
PANEL2	**Memory Treg**	**13.4**	**6.5**	**7 (0.5, 13.4)**	**0.0348**	**0.283**
PANEL2	**Naive Treg**	**17.4**	**3.5**	**13.8 (2.9, 24.8)**	**0.0146**	**0.283**
PANEL3	**B cells**	**59.3**	**10.8**	**48.5 (0.9, 96.2)**	**0.0463**	**0.283**
PANEL3	DN B cells	51.9	8.5	43.4 (-7, 93.8)	0.0877	0.297
PANEL3	Naive B cells	53.4	7.1	46.3 (-1.9, 94.5)	0.059	0.283
PANEL3	Transitional B cells	69.8	8.2	61.6 (-6.1, 129.3)	0.0722	0.283
PANEL3	Plasmablasts	52.3	40	12.3 (-11, 35.7)	0.292	0.485
PANEL3	Switched Memory B cells	50.4	12.4	38 (-13.8, 89.8)	0.142	0.345
PANEL3	**Non Switched Memory B cells**	**78.3**	**11.6**	**66.7 (8.5, 124.9)**	**0.0268**	**0.283**
PANEL3	*Plasma cells*	*86.9*	*54*	*32.9 (-13.9, 79.8)*	*0.162*	*0.359*
PANEL4	DCs	27.5	16	11.5 (-0.6, 23.6)	0.0621	0.283
PANEL4	Myeloid DCs	31.6	20.9	10.7 (-2.3, 23.8)	0.103	0.309
PANEL4	Plasmacytoid DCs	5.7	4.8	0.9 (-3.2, 5)	0.652	0.727
PANEL4	NK cells	4.7	2.3	2.4 (-1.6, 6.4)	0.222	0.419
PANEL4	CD16+ NK cells	4.9	2.4	2.5 (-1.1, 6.1)	0.169	0.359
PANEL4	CD16- NK cells	6.2	2.2	4 (-3.3, 11.3)	0.267	0.47
PANEL4	Monocytes	70.8	53.2	17.6 (-9.8, 45)	0.2	0.408
PANEL4	Classical Monocytes	70.7	53.7	17 (-10.5, 44.5)	0.217	0.419
PANEL4	**Non-classical Monocytes**	**80.6**	**47.7**	**32.9 (0.7, 65.1)**	**0.0455**	**0.283**
PANEL4	NKT cells	3.4	3.2	0.3 (-2.4, 2.9)	0.835	0.852

Notes: colours in the table: 1) purple indicates a significant difference of OX40L between SLE and controls with a p-value ≤ 0.05 & q-value > 0.05; 2) grey and italic indicates the cell subpopulation is excluded due to the limited number of cell counts.

a.



b.

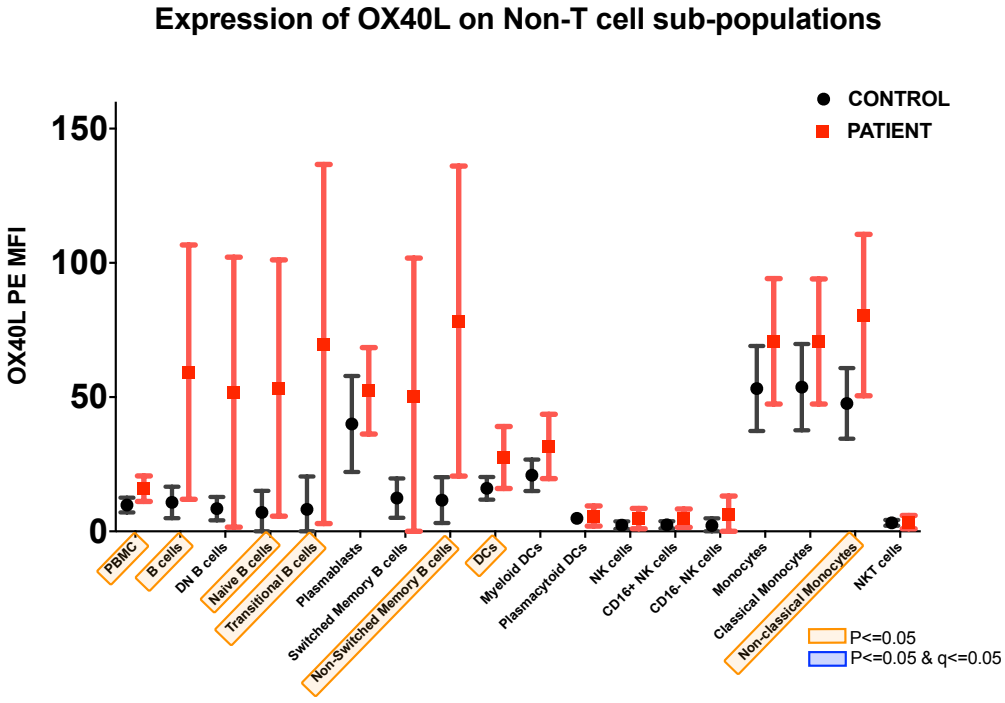


Figure 6- 10**. Protein expression of OX40L across all cell types identified.**
The x-axis shows the cell types and the y-axis shows the protein level (MFI) of OX40L. For the protein level of OX40L on the corresponding cell types, mean and 95%CI is showed. (a) T cell subset; (b) B cells, monocytes, NK cells and dendritic cell.

6.5 IKZF1 protein expression

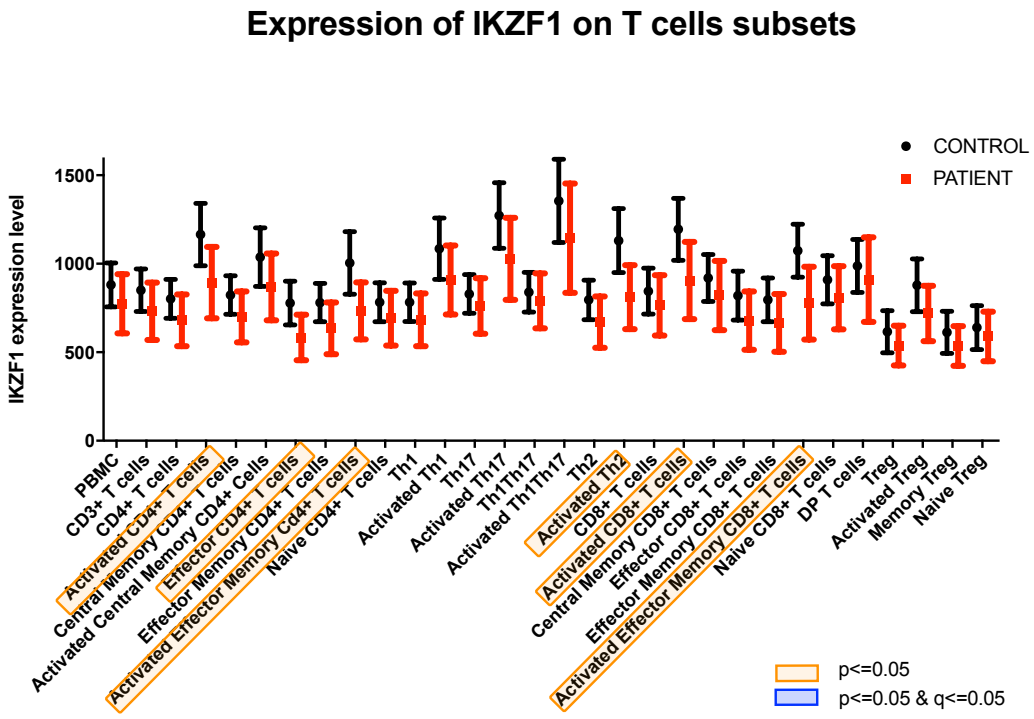
In general, SLE patients tended to have less IKZF1 protein levels across T cell subsets than in healthy controls (Figure 6-11 and Table 6-5). In the T cell panel, IKZF1 was significantly decreased in SLE patients for the following subsets: Activated CD4+ T cells (-272 (-531, -13), P = 0.0401), Effector CD4+ T cells (-195 (-367, -22), P = 0.028), Activated Effector memory CD4+ T cells (-271 (-502, -40), P = 0.023), Activated Th2 cells (-319 (-567, -72); P = 0.013), Activated CD8+ T cells (-289 (-561, -18), P = 0.037), and Activated Effector memory CD4+ T cells (-296 (-543, -49), P = 0.020), though all the significance level failed to pass the multiple testing burden ($q > 0.05$). No statistically significant differences were observed in the IKZF1 expression between SLE patients and healthy controls in B cell subsets, monocytes, NK cells, or DCs.

Table 6- 5. Average level of IKZF1 on immune subpopulations in SLE and controls.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	PBMC	774	881	-106 (-308, 96)	0.292	0.581
PANEL1	CD3 T cells	731	850	-119 (-314, 76)	0.225	0.522
PANEL1	CD4+ T cells	680	802	-121 (-299, 56)	0.175	0.454
PANEL1	**Activated CD4+ T cells**	**893**	**1165**	**-272 (-531, -13)**	**0.0401**	**0.256**
PANEL1	Central Memory CD4+ T cells	700	823	-124 (-299, 52)	0.162	0.454
PANEL1	Activated Central Memory CD4+ T cells	869	1038	-168 (-411, 74)	0.168	0.454
PANEL1	**Effector CD4+ T cells**	**583**	**778**	**-195 (-367, -22)**	**0.028**	**0.256**
PANEL1	*Activated Effector CD4+ T cells*	*776*	*1261*	*-484 (-791, -178)*	*0.00284*	*0.145*
PANEL1	Effector Memory CD4+ T cells	635	781	-146 (-322, 31)	0.103	0.454
PANEL1	**Activated Effector Memory CD4+ T cells**	**734**	**1005**	**-271 (-502, -40)**	**0.0229**	**0.256**
PANEL1	Naive CD4+ T cells	692	782	-90 (-274, 94)	0.328	0.581
PANEL1	Th1	683	783	-99 (-278, 80)	0.269	0.572
PANEL1	Activated Th1	908	1085	-177 (-429, 75)	0.164	0.454
PANEL1	Th17	761	829	-68 (-254, 118)	0.461	0.619
PANEL1	Activated Th17	1028	1272	-244 (-531, 43)	0.0928	0.454
PANEL1	Th1Th17	791	839	-48 (-234, 138)	0.602	0.663
PANEL1	Activated Th1Th17	1145	1355	-210 (-586, 166)	0.264	0.572
PANEL1	Th2	670	796	-126 (-303, 52)	0.16	0.454
PANEL1	**Activated Th2**	**811**	**1131**	**-319 (-567, -72)**	**0.0128**	**0.256**
PANEL1	CD8+ T cells	765	845	-80 (-288, 127)	0.436	0.619
PANEL1	**Activated CD8+ T cells**	**905**	**1195**	**-289 (-561, -18)**	**0.0371**	**0.256**
PANEL1	Central Memory CD8+ T cells	821	920	-99 (-329, 131)	0.388	0.586
PANEL1	*Activated Central Memory CD8+ T cells*	*1071*	*1456*	*-385 (-735, -35)*	*0.0318*	*0.256*
PANEL1	Effector CD8+ T cells	679	820	-141 (-349, 67)	0.178	0.454
PANEL1	*Activated Effector CD8+ T cells*	*842*	*1080*	*-238 (-586, 111)*	*0.175*	*0.454*
PANEL1	Effector Memory CD8+ T cells	666	796	-130 (-328, 68)	0.192	0.466
PANEL1	**Activated Effector Memory CD8+ T cells**	**778**	**1074**	**-296 (-543, -49)**	**0.0202**	**0.256**
PANEL1	Naive CD8+ T cells	808	909	-101 (-318, 117)	0.354	0.582
PANEL1	DP T cells	911	988	-77 (-351, 198)	0.574	0.657
PANEL2	Treg	538	616	-78 (-236, 80)	0.326	0.581
PANEL2	Activated Treg	719	878	-159 (-368, 50)	0.131	0.454
PANEL2	Memory Treg	535	613	-78 (-235, 80)	0.326	0.581
PANEL2	Naive Treg	590	639	-50 (-231, 131)	0.58	0.657
PANEL3	B cells	753	674	79 (-133, 291)	0.451	0.619
PANEL3	DN B cells	856	764	92 (-123, 307)	0.391	0.586
PANEL3	Naive B cells	604	580	24 (-152, 201)	0.78	0.829
PANEL3	Transitional B cells	714	713	2 (-169, 173)	0.983	0.983
PANEL3	Plasmablasts	1115	1348	-233 (-523, 56)	0.111	0.454
PANEL3	Switched Memory B cells	892	876	15 (-191, 222)	0.88	0.91
PANEL3	Non Switched Memory B cells	750	764	-14 (-222, 194)	0.892	0.91
PANEL3	*Plasma cells*	*1004*	*1166*	*-163 (-687, 362)*	*0.53*	*0.644*
PANEL4	DCs	1390	1552	-162 (-496, 171)	0.331	0.581
PANEL4	Myeloid DCs	1364	1446	-82 (-405, 241)	0.611	0.663
PANEL4	Plasmacytoid DCs	1888	1735	154 (-282, 589)	0.477	0.624
PANEL4	NK cells	794	734	60 (-117, 237)	0.496	0.632
PANEL4	CD16+ NK cells	787	734	53 (-122, 228)	0.545	0.646
PANEL4	CD16- NK cells	806	740	66 (-114, 245)	0.461	0.619
PANEL4	Monocytes	1043	1165	-122 (-398, 154)	0.378	0.586
PANEL4	Classical Monocytes	1010	1138	-128 (-397, 141)	0.342	0.581
PANEL4	Non-classical Monocytes	1550	1821	-271 (-671, 129)	0.178	0.454
PANEL4	NKT cells	638	689	-51 (-210, 109)	0.522	0.644

Notes: colours in the table: 1) purple indicates a significant difference of IKZF1 between SLE and controls with a p-value ≤ 0.05 & q-value > 0.05; 2) grey and italic indicates the cell subpopulation is excluded due to the limited number of cell counts.

a.



b.

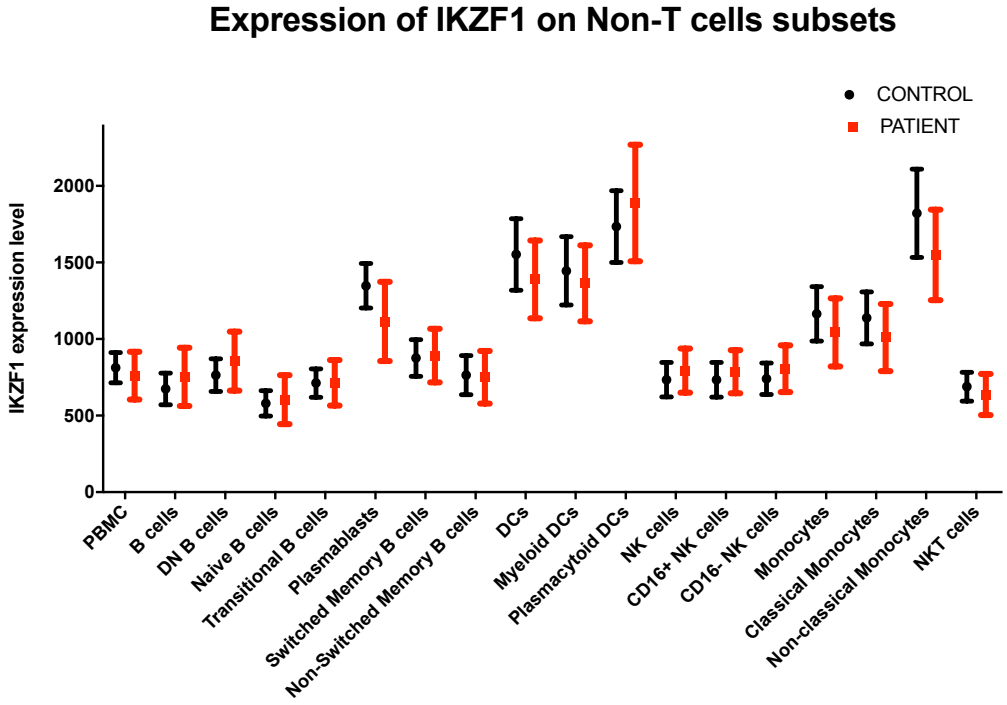


Figure 6- 11**. Protein expression of IKZF1 across all cell types identified.**

The x-axis shows the cell types and the y-axis shows the protein level (MFI) of IKZF1. For the protein level of IKZF1 on the corresponding cell types, mean and 95%CI is showed.

6.6 IKZF2 protein expression

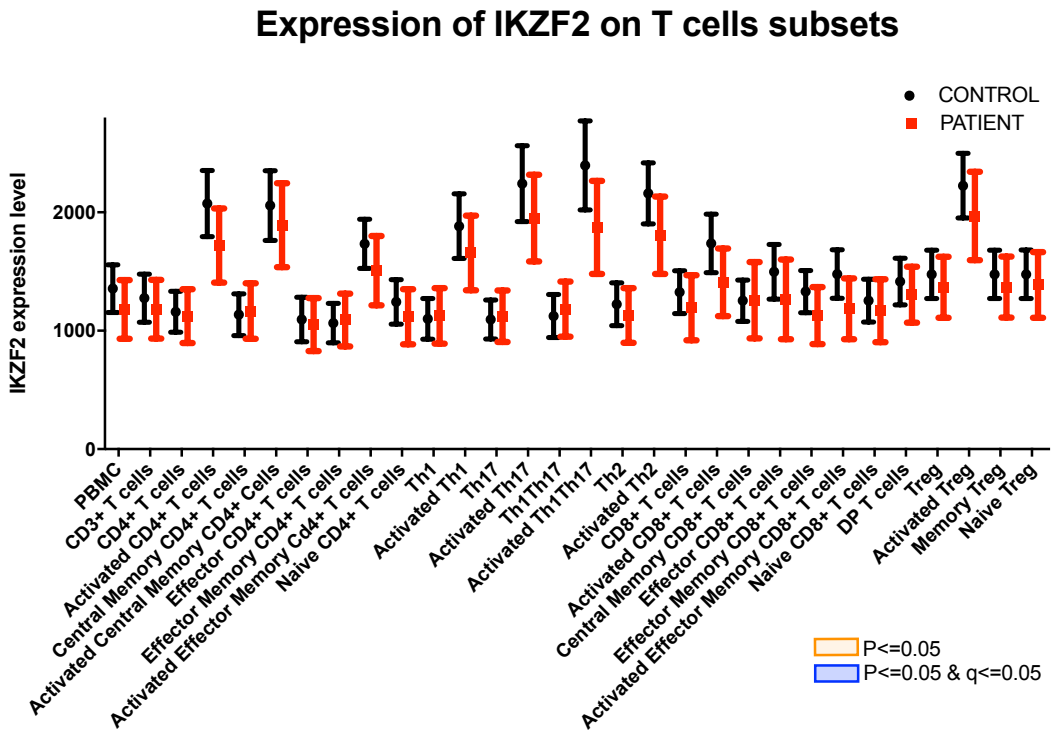
No significant differences were observed in terms of IKZF2 expression between SLE patients and healthy controls across all T cell subsets. However, a higher level of IKZF2 was found in activated CD4+ T cell subsets than in non-activated ones in both patients and controls as shown in Figure 6-12 and Table 6-6. For example, protein level of IKZF2 was higher in activated Th1 than in Th1 cells, and higher in activated regulatory T cells than in naïve regulatory T cells. In the non-T cells panel, NK cells (131 (-272, 534), P = 0.51) showed a trend of increasing in IKZF2 in patients, while monocytes (-160 (-429, 109); P = 0.23) and dendritic cells (-122 (-464, 220), P = 0.47) showed a trend of decrease in patients than in controls. No clear difference was observed between patients and controls in the B cell subsets. Overall, differences of IKZF2 protein level between SLE patients and controls failed to reach the statistically significance across all immune subsets under detection.

Table 6- 6. Average level of IKZF2 on immune subpopulations in SLE and controls.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	PBMC	1180	1355	-176 (-485, 134)	0.258	0.848
PANEL1	CD3 T cells	1182	1275	-93 (-404, 218)	0.548	0.942
PANEL1	CD4+ T cells	1124	1160	-37 (-314, 241)	0.791	0.942
PANEL1	Activated CD4+ T cells	1720	2074	-354 (-761, 53)	0.0864	0.848
PANEL1	Central Memory CD4+ T cells	1167	1135	31 (-253, 316)	0.824	0.942
PANEL1	Activated Central Memory CD4+ T cells	1892	2057	-165 (-612, 281)	0.458	0.934
PANEL1	Effector CD4+ T cells	1052	1095	-43 (-325, 240)	0.761	0.942
PANEL1	*Activated Effector CD4+ T cells*	*1348*	*1572*	*-224 (-710, 262)*	*0.357*	*0.934*
PANEL1	Effector Memory CD4+ T cells	1090	1064	26 (-243, 296)	0.844	0.942
PANEL1	Activated Effector Memory CD4+ T cells	1507	1734	-226 (-575, 122)	0.195	0.848
PANEL1	Naive CD4+ T cells	1118	1243	-125 (-415, 165)	0.387	0.934
PANEL1	Th1	1125	1100	24 (-258, 307)	0.863	0.942
PANEL1	Activated Th1	1657	1883	-226 (-630, 179)	0.266	0.848
PANEL1	Th17	1122	1094	28 (-237, 293)	0.831	0.942
PANEL1	Activated Th17	1951	2242	-291 (-763, 181)	0.219	0.848
PANEL1	Th1Th17	1182	1124	58 (-229, 345)	0.684	0.942
PANEL1	Activated Th1Th17	1873	2396	-523 (-1050, 3)	0.0511	0.848
PANEL1	Th2	1129	1223	-94 (-379, 190)	0.506	0.934
PANEL1	Activated Th2	1806	2160	-354 (-759, 51)	0.085	0.848
PANEL1	CD8+ T cells	1194	1326	-132 (-451, 188)	0.409	0.934
PANEL1	Activated CD8+ T cells	1409	1738	-329 (-695, 37)	0.0771	0.848
PANEL1	Central Memory CD8+ T cells	1258	1253	5 (-354, 363)	0.979	0.979
PANEL1	*Activated Central Memory CD8+ T cells*	*1603*	*1915*	*-312 (-747, 123)*	*0.155*	*0.848*
PANEL1	Effector CD8+ T cells	1264	1497	-233 (-630, 164)	0.241	0.848
PANEL1	*Activated Effector CD8+ T cells*	*1352*	*1562*	*-210 (-671, 251)*	*0.362*	*0.934*
PANEL1	Effector Memory CD8+ T cells	1128	1330	-202 (-494, 89)	0.167	0.848
PANEL1	Activated Effector Memory CD8+ T cells	1185	1479	-294 (-612, 25)	0.0697	0.848
PANEL1	Naive CD8+ T cells	1169	1254	-84 (-397, 228)	0.586	0.942
PANEL1	DP T cells	1419	1455	-36 (-422, 350)	0.851	0.942
PANEL2	Treg	1367	1475	-108 (-427, 210)	0.494	0.934
PANEL2	Activated Treg	1969	2225	-256 (-705, 193)	0.255	0.848
PANEL2	Memory Treg	1369	1476	-108 (-427, 211)	0.498	0.934
PANEL2	Naive Treg	1387	1477	-90 (-425, 245)	0.589	0.942
PANEL3	B cells	994	957	38 (-227, 302)	0.774	0.942
PANEL3	DN B cells	907	882	25 (-209, 260)	0.827	0.942
PANEL3	Naive B cells	932	903	29 (-249, 307)	0.832	0.942
PANEL3	Transitional B cells	1206	1184	22 (-266, 309)	0.88	0.942
PANEL3	Plasmablasts	754	714	40 (-212, 292)	0.747	0.942
PANEL3	Switched Memory B cells	1013	1031	-19 (-281, 244)	0.887	0.942
PANEL3	Non Switched Memory B cells	1082	1059	23 (-264, 310)	0.873	0.942
PANEL3	*Plasma cells*	*653*	*644*	*8 (-423, 439)*	*0.969*	*0.979*
PANEL4	DCs	1287	1408	-122 (-464, 220)	0.475	0.934
PANEL4	Myeloid DCs	1359	1516	-158 (-529, 214)	0.396	0.934
PANEL4	Plasmacytoid DCs	1122	1130	-8 (-314, 299)	0.96	0.979
PANEL4	NK cells	1700	1569	131 (-272, 534)	0.513	0.934
PANEL4	CD16+ NK cells	1786	1608	178 (-242, 598)	0.395	0.934
PANEL4	CD16- NK cells	1264	1223	41 (-258, 340)	0.783	0.942
PANEL4	Monocytes	1025	1185	-160 (-429, 109)	0.235	0.848
PANEL4	Classical Monocytes	1001	1166	-165 (-429, 99)	0.212	0.848
PANEL4	Non-classical Monocytes	1572	1836	-264 (-733, 205)	0.261	0.848
PANEL4	NKT cells	1179	1266	-87 (-417, 243)	0.596	0.942

Notes: text in grey and italic indicates the cell subpopulation is excluded due to the limited number of cell counts.

a.



b.

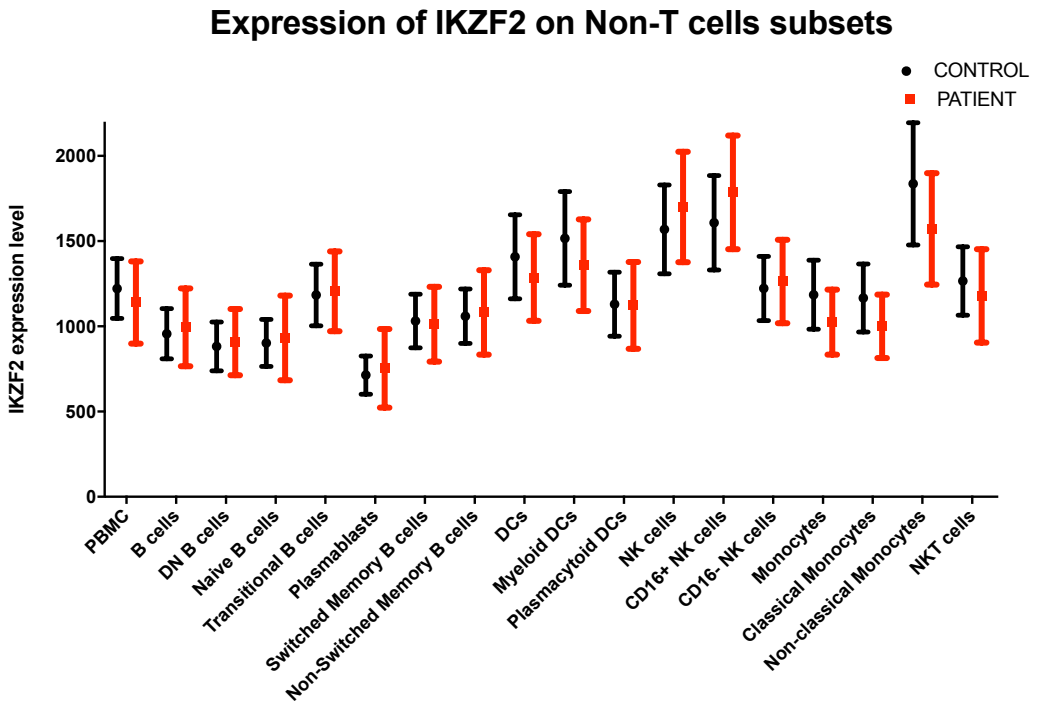


Figure 6- 12**. Protein expression of IKZF2 across all immune subsets.**

The x-axis shows the cell types and the y-axis shows the protein level (MFI) of IKZF2. For the protein level of IKZF2 on the corresponding cell types, mean and 95%CI is showed.

6.7 IKZF3 protein expression

In general, the expression level of IKZF3 was decreased in SLE patients compared to healthy controls in T cell subsets (Figure 6-13 and Table 6-7). Similar to IKZF2, IKZF3 the protein level is increased in activated CD4+ T cells subsets comparing to the non-activated ones. Specifically, the differences between activated CD4+ T cells were exaggerated after activation, i.e., the differences of IKZF3 protein level between SLE and controls in Activated CD4+ T cells (-165 (-323, -6), P = 0.042), Activated Central Memory CD4+ T cells (-154 (-285, -23), P = 0.0225), Activated Th1 (-160 (-306, -13), P = 0.0339), Activated Th17 (-198 (-370, -26), P = 0.0249), Activated Th2 (-146 (-282, -10), P = 0.0356) and Activated Regulatory T cells (-122 (-229, -15), P = 0.026) were statistically significant though failed to pass the multiple testing burdens (q > 0.05).

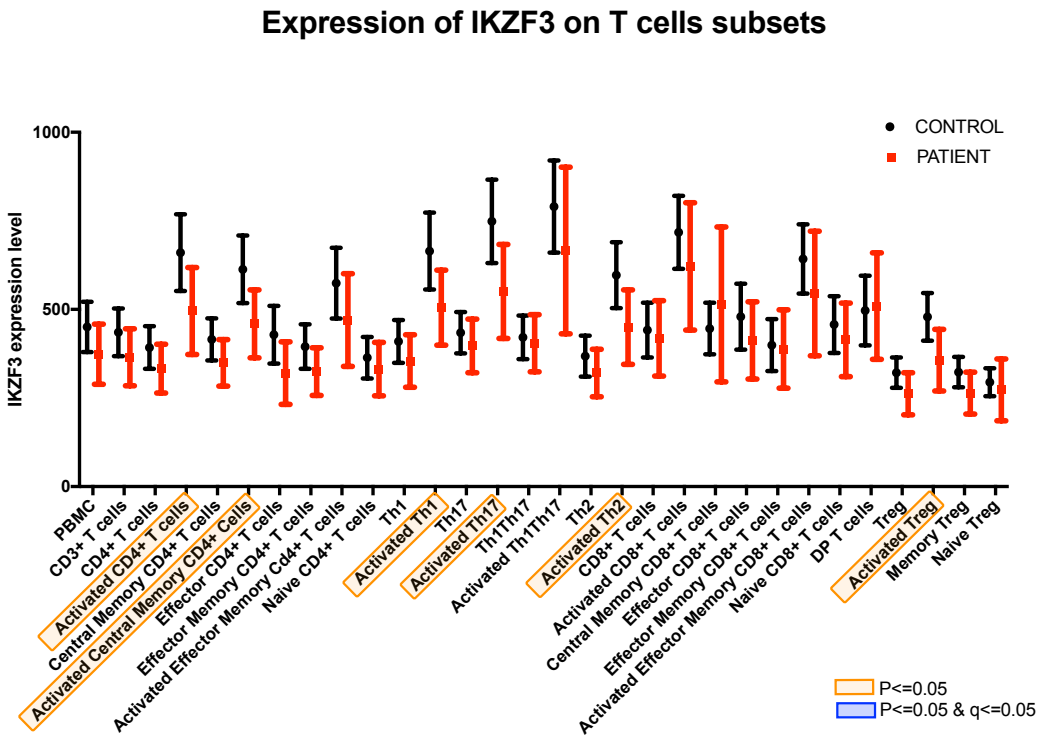
The protein level of IKZF3 in the non-T cells subsets were similar between SLE and controls with modest differences in certain cell subsets. Notably, the expression level of IKZF3 peaked in plasmablasts (MFI ranging from 800 ~ 1000) in both patients and controls. Specifically, IKZF3 protein level was increased in NK cells while decreased in monocytes in SLE patients when comparing to healthy controls.

Table 6- 7. Average level of IKZF3 on immune subpopulations in SLE and controls.

PANEL	Cell Type	SLE Mean	Control Mean	Mean difference (95% CI)	p-value	q-value
PANEL1	PBMC	374	451	-77 (-184, 30)	0.155	0.549
PANEL1	CD3 T cells	365	436	-70 (-172, 31)	0.169	0.549
PANEL1	CD4+ T cells	333	392	-59 (-148, 29)	0.183	0.549
PANEL1	**Activated CD4+ T cells**	**496**	**660**	**-165 (-323, -6)**	**0.042**	**0.306**
PANEL1	Central Memory CD4+ T cells	349	415	-66 (-152, 20)	0.127	0.498
PANEL1	**Activated Central Memory CD4+ T cells**	**459**	**613**	**-154 (-285, -23)**	**0.0225**	**0.303**
PANEL1	Effector CD4+ T cells	320	428	-108 (-225, 8)	0.0682	0.386
PANEL1	*Activated Effector CD4+ T cells*	*497*	*837*	*-340 (-548, -133)*	*0.00202*	*0.103*
PANEL1	Effector Memory CD4+ T cells	324	395	-71 (-160, 19)	0.117	0.497
PANEL1	Activated Effector Memory CD4+ T cells	470	574	-104 (-264, 56)	0.196	0.555
PANEL1	Naive CD4+ T cells	332	364	-32 (-125, 61)	0.487	0.792
PANEL1	Th1	354	410	-55 (-148, 38)	0.235	0.599
PANEL1	**Activated Th1**	**505**	**665**	**-160 (-306, -13)**	**0.0339**	**0.303**
PANEL1	Th17	397	434	-37 (-130, 56)	0.421	0.74
PANEL1	**Activated Th17**	**551**	**749**	**-198 (-370, -26)**	**0.0249**	**0.303**
PANEL1	Th1Th17	405	421	-17 (-115, 82)	0.732	0.879
PANEL1	Activated Th1Th17	666	790	-124 (-386, 139)	0.344	0.675
PANEL1	Th2	321	368	-48 (-134, 38)	0.269	0.635
PANEL1	**Activated Th2**	**450**	**597**	**-146 (-282, -10)**	**0.0356**	**0.303**
PANEL1	CD8+ T cells	418	442	-23 (-151, 104)	0.712	0.879
PANEL1	Activated CD8+ T cells	621	718	-96 (-299, 106)	0.338	0.675
PANEL1	Central Memory CD8+ T cells	514	446	68 (-159, 296)	0.542	0.838
PANEL1	*Activated Central Memory CD8+ T cells*	*745*	*866*	*-121 (-440, 198)*	*0.446*	*0.758*
PANEL1	Effector CD8+ T cells	412	480	-67 (-206, 72)	0.334	0.675
PANEL1	*Activated Effector CD8+ T cells*	*581*	*629*	*-47 (-289, 194)*	*0.693*	*0.879*
PANEL1	Effector Memory CD8+ T cells	388	399	-11 (-140, 118)	0.864	0.958
PANEL1	Activated Effector Memory CD8+ T cells	545	642	-98 (-294, 99)	0.318	0.675
PANEL1	Naive CD8+ T cells	414	457	-43 (-170, 84)	0.497	0.792
PANEL1	DP T cells	509	497	13 (-162, 187)	0.885	0.96
PANEL2	Treg	262	322	-60 (-131, 11)	0.0977	0.453
PANEL2	**Activated Treg**	**357**	**479**	**-122 (-229, -15)**	**0.026**	**0.303**
PANEL2	Memory Treg	264	323	-59 (-130, 11)	0.097	0.453
PANEL2	Naive Treg	273	295	-22 (-116, 72)	0.637	0.855
PANEL3	B cells	596	532	64 (-86, 215)	0.391	0.712
PANEL3	DN B cells	642	602	40 (-114, 193)	0.603	0.839
PANEL3	Naive B cells	461	442	19 (-95, 133)	0.741	0.879
PANEL3	Transitional B cells	591	505	86 (-40, 212)	0.174	0.549
PANEL3	Plasmablasts	831	924	-93 (-299, 113)	0.365	0.689
PANEL3	Switched Memory B cells	726	750	-24 (-192, 144)	0.773	0.883
PANEL3	Non Switched Memory B cells	587	620	-32 (-179, 114)	0.659	0.862
PANEL3	*Plasma cells*	*759*	*751*	*8 (-468, 484)*	*0.972*	*0.972*
PANEL4	DCs	395	399	-4 (-110, 102)	0.941	0.972
PANEL4	Myeloid DCs	390	375	15 (-91, 120)	0.779	0.883
PANEL4	Plasmacytoid DCs	398	396	3 (-91, 97)	0.954	0.972
PANEL4	NK cells	501	447	54 (-45, 154)	0.274	0.635
PANEL4	CD16+ NK cells	516	455	62 (-39, 163)	0.222	0.596
PANEL4	CD16- NK cells	364	287	77 (-1, 155)	0.053	0.338
PANEL4	Monocytes	337	365	-27 (-134, 79)	0.609	0.839
PANEL4	Classical Monocytes	326	355	-29 (-133, 76)	0.584	0.839
PANEL4	Non-classical Monocytes	454	493	-39 (-179, 102)	0.579	0.839
PANEL4	NKT cells	397	400	-3 (-93, 87)	0.944	0.972

Notes: colours in the table: 1) purple indicates a significant difference of IKZF3 between SLE and controls with a p-value ≤ 0.05 & q-value > 0.05; 2) grey and italic indicates the cell subpopulation is excluded due to the limited number of cell counts.

a.



b.

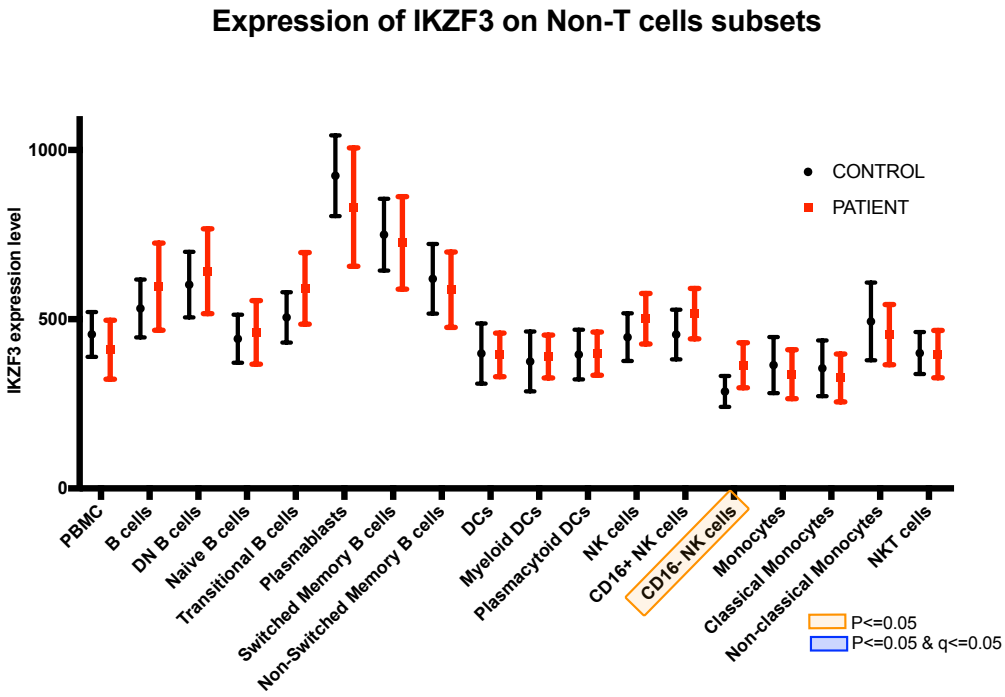


Figure 6- 13. **Protein expression of IKZF3 across all cell types identified.**
The x-axis shows the cell types and the y-axis shows the protein level (MFI) of IKZF3. For the protein level of IKZF3 on the corresponding cell types, mean and 95%CI is showed.

6.8 Correlation between protein expression and SLEDAI

Only SLEDAI for 15 patients were available, see Chapter 2. Methods, Table 2-4. In order to test if the protein level of target molecules (OX40L and IKZF1/2/3) correlated with the disease activity, I performed a Spearm correlation test between SLEDAI and the protein level of target molecules for all cell types identified. Only the expression of OX40L on Th17 cells showed significant correlation with SLEDAI (P = 0.029) (Figure 6-14a). Protein level of IKZF3 on B cells (P = 0.089) and plasmablasts (P = 0.097) showed suggestive correlation with SLEDAI, with higher level of IKZF3 correlated with more active disease status (Figure 6-14b). Similarly, increasing expression of IKZF2 in monocytes and DCs were positively associated with SLEDAI. Interestingly, only expression of IKZF2 on non-classical monocytes (P = 0.069) was observed to be correlated with the activity of SLE, but not classical monocytes (P = 0.308), suggesting the differential roles of IKZF2 on monocytes subpopulations (Figure 6-14c). While for the DCs (Figure 6-14d), only IKZF2 on myeloid DCs (P = 0.099) showed slightly correlation with SLEDAI, but not plasmacytoid DCs (P= 0.422). No statistically significant correlations were observed for the protein level of IKZF2 with SLEDAI.

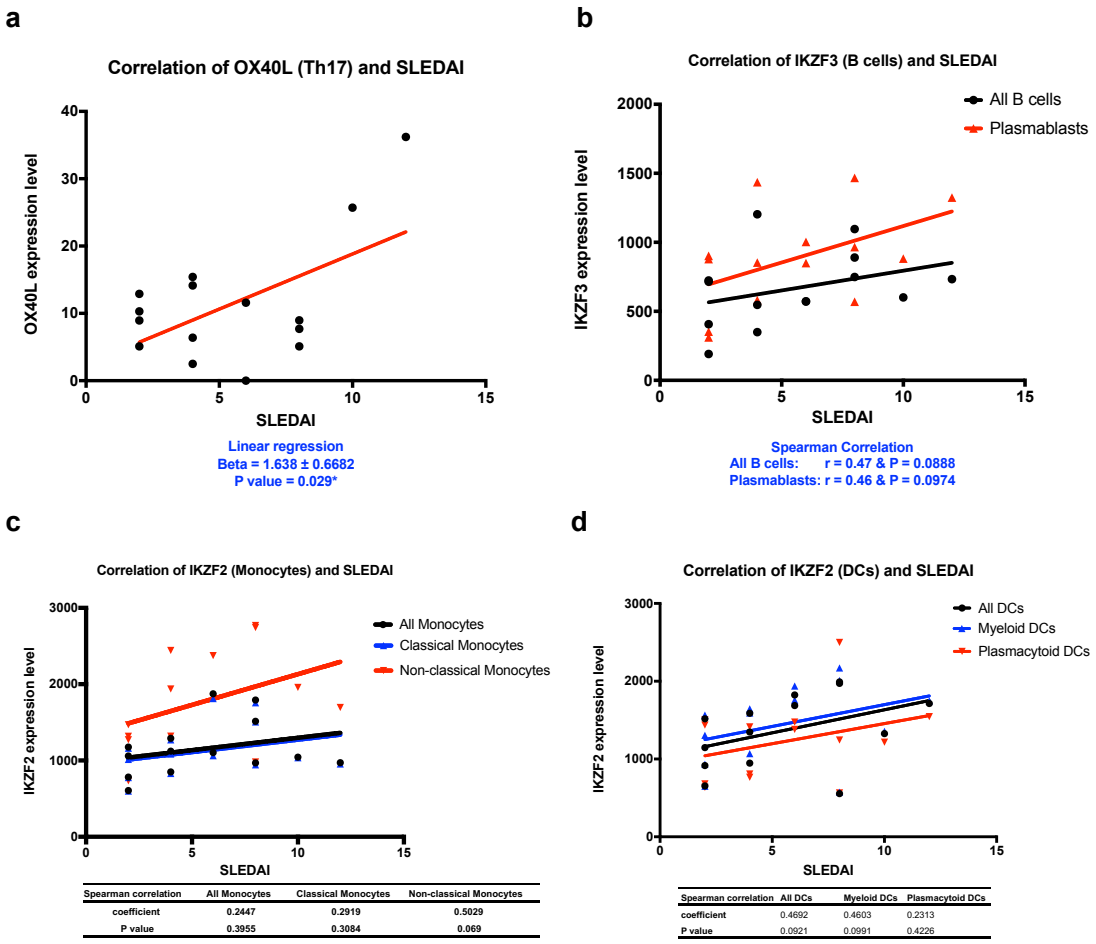


Figure 6- 14. **Correlation between protein expression and SLEDAI.**

For all the scatter plots, SLEDAI is plotted on the x-axis and the protein level of target molecule is showed on the y-axis. The lines within the scatter plots are their corresponding linear regression lines. The summary statistics are showed under each plot. a) Correlation between OX40L on Th17 cells and SLEDAI. b) For correlation between IKZF3 and SLEDAI, red and black represent IKZF3 in plasmablasts and B cells, respectively. c) For correlation between IKZF2 in monocytes and SLEDA, black, blue and red represent IKZF2 in all monocytes, classical monocytes, and non-classical monocytes, respectively. d) For correlation between IKZF2 in DCs and SLEDA, black, blue and red represent IKZF2 in all DCs, Myeloid DCs, and plasmacytoid DCs, respectively.

6.9 Discussion

Genetic studies have identified robust associations of *IKZF1, IKZF2*, *IKZF3* and *OX40L* with SLE (Chapter 1. Introduction, Figure 1-8) [43]. Functional studies on the Ikaros transcription factor family [114, 202-205] and OX40L [201] have shown that dysfunctions of these factors will lead to autoimmune and inflammatory disorders. Since PBMC (peripheral blood mononuclear cell) has a heterogeneous composition, comprising a variety of different immune cells with discrete functions, measuring protein level of target molecules in each immune subpopulation provides clues to infer their potential functions. Hence, I measured the protein products of the target genes across a range of immune subsets in PBMC derived from 22 SLE patients and 20 healthy controls. The panel design for detecting immune subpopulations of T, B and NK cells, monocytes, and DCs is mainly derived from a set of defined eight-colour antibody cocktails proposed by the Human Immunophenotyping Consortium (HIPC) [100], but with some modifications as we want to measure the protein level of four target genes simultaneously. That is to say, we used the cellular markers as proposed by HIPC (Chapter 2. Methods, Figure 2-6to distinguish the immune sub-populations, but the antibody clones and fluorochromes for each cellular marker were changed when needed. Though Finak et al [206] suggested that the evaluation of helper T cell panel (Th1/2/17 panel) by the HIPC [100] is too variable to be reliable, the data quality of this panel in my study was relatively good as the modified antibody-conjugates for these cell subsets could distinguish them clearly (Figure 6-1).

By comparing the absolute number of cells as well as the cell proportions of immune subsets between SLE and controls, I observed that either the absolute number or cell proportions of immune cells from adaptive immune system were increased in SLE patients when comparing to healthy controls, including naïve CD8+ T cells, Th2 cells, activated CD4+ T cells, activated Th1 cells, activated Th2 cells, and B cells. While immune cells derived from the innate immune system were generally observed to be decrease in SLE patients, such as plasmacytoid DCs and CD16+ NK cells (Figure 6-6/8). In addition, modest negative correlations were observed between proportions of immune subsets, such as NK and B cell, and SLEDAI (Figure 6-7/9), suggesting that patients with more active SLE tended to have more severe leukopenia.

For the first time, this study shows the protein expression pattern of OX40L and the Ikaros transcription factor family (IKZF1/2/3) across a range of immune cells (Figure 6-10/11/12/13). In general, protein levels of OX40L is increased while IKZF1, IKZF2 and IKZF3 are decreased in SLE patients. GWAS and family-based studies [41, 45, 128] have underlined the importance of *OX40L* as a SLE susceptibility gene, with an increased expression of *OX40L* in individuals carrying the disease risk allele according to the eQTL results found in the FHS whole blood dataset [85]. Consistently, the protein level of OX40L on B cells was significantly increased in SLE patients in my study. The higher level of OX40L could augment the B-cell hyperactivity and the resulting autoantibodies and immune complexes may lead to the development of autoimmune diseases. Studies by Jacquemin et al [207] and Cortini et al [201] both showed that OX40L-OX40 axis plays a major role in lupus pathogenesis by promoting the responses of T follicular helper (T_{FH}) cells. Cortini et al's study applying a spontaneous congenic lupus murine model shows that OX40L on B cells facilitates the development of T_{FH}. However, no correlation between blood B cells expressing OX40L and T_{FH} cells was observed in Jacquemin et al's study on SLE patients, which could be due to the compartmentalisation of activated B cells expressing OX40L in the secondary lymphoid organs rather than an evidence of their lack of involvement in the development of pathogenic T_{FH} cells in SLE as suggested by Cortini et al.

The Ikaros transcription factor family, including IKZF1, IKZF2 and IKZF3 are genetically associated with a variety of human autoimmune diseases [43, 208-217]. eQTL studies based on the exon-level and junction level suggest that non-coding aberrations disrupt regulation of expression of these genes [80, 184, 218]. Though not all the differences between cases and controls in my study are statistically significant, the expression patterns of all three Ikaros family members are quite similar – their protein products are decreased in patients comparing to controls in the T cells panel. While in NK cells, all there IKZFs are slightly higher in SLE patients than in controls, suggesting cell specificity should be taken into account when studying target genes' functions.

IKZF1 (Ikaros) is important to limit autoimmunity and its down regulation is associated with autoimmune disease [219]. Consistently, I observed decreased Ikaros in SLE across a variety of immune cell subsets. In particular cell types, Ikaros was significantly lower in patients than that in controls, including effector CD4+ T cells, activated CD4+ T cells, activated Th2 cells, and activated CD8+ T cells (p-value $\leq$ 0.05) (Figure 6-11 and Table 6-5). Further studies to

investigate the underlying roles of Ikaros in activated T cell may provide insights to the pathogenesis of SLE. IKZF1 acts as both a transcriptional suppressor and activator depending on the co-factors it interacts [220] . Studies in the murine system [221-223] show that Ikaros co-factors are over-expressing in non-lymphoid cell lines with suppressive functions. While Ikaros may also function as a transcriptional activator possibly by recruiting SWI/SNF or promoting transcriptional elongation [224, 225]. Moreover, Ikaros is found to control multiple levels of B cell lymphopoiesis and function, which is required for earliest B cell progenitors [226], efficient pro-B to pre-B transition [227, 228], repression of non-B cell fates [228], and promoting Igh rearrangement and Rag expression [228, 229]. I observed a higher level of Ikaros in DCs and monocytes compared to T cells and B cells in both patients and controls, supporting the suppressive roles of Ikaros in the non-lymphoid cell lines.

IKZF2 (Helios) is expressed in early hematopoietic progenitors of the bone marrow and Treg cells [112], and regulates Treg associated genes, such as FoxP3 [230, 231]. Functional analysis using ChIP and knocking-down techniques found that suppression of IKZF2 in Treg significantly attenuates their suppressive function [113]. Although no significant differences of IKZF2 between SLE and controls have been found, modest correlations between SLEDAI and IKZF2 were observed in non-classical monocytes and myeloid DCs (Figure 6-12), suggesting the perturbation of IKZF2 in specific immune subsets may correlate with the subtle progression in disease course. Consistently, previous studies [114, 232] reported that IKZF2+ Treg were significantly increased in SLE patients and expanded in active SLE, supporting the concept that IKZF2 suppress the function of Treg, leading to the dysfunction of immune suppressive effect of Treg and resulting in an increase of autoimmune response.

IKZF3 (Aiolos) is a transcription factor helps regulate transition of Pro-B cells and Pre-B cells and is critical in differentiation of plasma cells and development of B-cell memory [233]. IKZF3 knock-out mice which lack B-cell immunological memory develop SLE-like disease as well as haematological malignancies [234]. Recent studies have reported that IKZF3 is overexpressed in SLE B cell subtype [235] and PBMCs [205]. Consistently, I show that the expression of IKZF3 on B cell subsets is slightly increased in SLE patients (Figure 6-13) and is positively correlated with SLEDAI (Figure 6-14), supporting the important role of IKZF3 in the development and progression of SLE. Intriguingly, the expression of IKZF3 is significantly decreased in SLE in activated CD4+ T cells and some of the sub-populations, i.e. activated Central Memory CD4+ T

cells, activated Th1, activated Th2, activated Th17 and activated Treg, suggesting a potential role of IKZF3 in activated T cells may be involved in SLE.

In summary, the comprehensive immunophenotyping of target molecules' protein expression to scrutinize the subtle differences across discrete immune subsets provides insights into the functions of the target molecules in the pathogenesis of SLE. In general, the expression of these molecules (OX40L, IKZF1, IKZF2, and IKZF4) is higher/lower in some cell types than others, indicating their involvements and functions in the immune system are likely to be cell specific. Understanding the cell-specific effects of these molecules involved in the immune system holds great promise for developing them as potential therapeutic targets in the manipulation of SLE. Therefore, further functional studies should focus on specific cell types, including those that have enriched expression of particular molecules (e.g., IKZF2 in Tregs; OX40L in B cells and APCs) and those that exhibit differential expression levels between cases and controls (e.g. IKZF3 in T cells and B cells), in order to better scrutinize these molecules' potential roles on their dominant cell types. Moreover, genotyping all these samples, including patients and controls, will provide valuable insights to bridge the gap between genetics and protein expression target molecules.

Chapter 7. Conclusions

In this thesis, I integrated genetics, epigenetics and gene expression to delineate the regulatory functions of SLE risk loci and their roles on the severity of SLE. I believe my findings are novel contributions to the understanding of SLE genetics.

I have demonstrated that eQTL mapping, incorporating co-localization analysis of genetic association across a variety of immune cell types is essential to detect the full array of disease associated eQTLs. By doing so in chapter 3, I have identified candidate-causal genes for almost half of the SLE risk loci. In particular, my data have pinpointed some candidate-causal genes acting in specific cell type or under certain context (resting or stimulation), which warrant further study into certain condition to delineate the molecular function of causal genes. Moreover, by mapping eQTLs across populations, I have identified candidate-causal eGenes which may be masked by strong LD within genomic regions when analysing in a single population. Though expression data generated from MicroArray are a good surrogate for measuring transcript at probe level, using RNA-seq to quantify transcript level at multiple resolutions, including gene-, exon-, splice-junction, and isoform, is very likely to advance our understanding of the implications of the genetic control of gene expression.

In Chapter 4, I expanded my analyses on the functional roles of disease associated loci with regulatory epigenomes across a variety of tissues and primary cells and showed that SLE risk loci were primarily enriched in regulatory regions within a range of immune cells. Furthermore, I pinpointed that a large proportion (66%) of SLE eQTLs were overlapped with DHS hotspots, denoting extensive local coordination of genetic influences on gene expression and epigenetics. With the development of NGS technologies, the resolution on measuring gene expression and epigenetics is largely improved and the power to detect QTLs (i.e., expression QTL and chromatin accessibility QTL) is much extended. Therefore, I emphasize that large gene expression and epigenome datasets for QTL studies should be generated across trait-associated cell types and under disease conditions. An alternative and most accessible approach maybe to perform

expression and chromatin assays (RNA-seq and ATAC-seq) in whole-blood and later decompose measurements into component cell types statistically.

As kidney involvement is strongly associated with disease severity of SLE, I chose this sub-phenotype to investigate its relationship with the SLE genetic risk factors in Chapter 5. Due to the lack of significant genetic associations in SLE Renal GWAS, I tested the hypothesis that the genetic risk loading of SLE may correlate with kidney involvement in a quantitative way and found that the genetic risk factors that influence the severity of SLE were through a quantitative way (multiple genes' contribution, GRS). The inclusion of SLE risk alleles into a GRS can modestly predict the renal involvement in SLE patients (LN), shows consistent discriminatory ability in independent cohorts, and is enhanced by the inclusion of age-of-onset into the model. Therefore, I suggest that the GRS will make a contribution to predicting LN in SLE patients in a clinical or research setting.

As part of the functional study of post-GWAS in Chapter 6, I investigated the protein expression of selected SLE-susceptibility gene products (the Ikaros family members and OX40L) in a range of immune cells, using multi-parameter flow cytometry. The results reveal some cellular specificity in gene expression in disease. In particular, IKZF3 expression on activated regulatory T cell and helper T cell subsets was decreased, while OX40L expression on B cell subsets was increased in SLE. This is a comprehensive immunephenotyping of target molecules to dissect their potential roles in SLE, which provides good evidence for further functional experiments to focus on target cell types, thereby leading to faster turnaround and lower costs in molecular functional studies.

References

1. Kaul, A., et al., *Systemic lupus erythematosus.* Nat Rev Dis Primers, 2016. **2**: p. 16039.
2. Feldman, C.H., et al., *Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000-2004.* Arthritis Rheum, 2013. **65**(3): p. 753-63.
3. Carter, E.E., S.G. Barr, and A.E. Clarke, *The global burden of SLE: prevalence, health disparities and socioeconomic impact.* Nat Rev Rheumatol, 2016. **12**(10): p. 605-20.
4. Nasonov, E., et al., *The prevalence and incidence of systemic lupus erythematosus (SLE) in selected cities from three Commonwealth of Independent States countries (the Russian Federation, Ukraine and Kazakhstan).* Lupus, 2014. **23**(2): p. 213-9.
5. Brinks, R., et al., *Age-specific and sex-specific incidence of systemic lupus erythematosus: an estimate from cross-sectional claims data of 2.3 million people in the German statutory health insurance 2002.* Lupus Sci Med, 2016. **3**(1): p. e000181.
6. Alamanos, Y., et al., *Epidemiology of systemic lupus erythematosus in northwest Greece 1982-2001.* J Rheumatol, 2003. **30**(4): p. 731-5.
7. Molokhia, M., et al., *Systemic lupus erythematosus in migrants from west Africa compared with Afro-Caribbean people in the UK.* Lancet, 2001. **357**(9266): p. 1414-5.
8. Pons-Estel, G.J., M.F. Ugarte-Gil, and G.S. Alarcon, *Epidemiology of systemic lupus erythematosus.* Expert Rev Clin Immunol, 2017. **13**(8): p. 799-814.
9. Block, S.R., et al., *Studies of twins with systemic lupus erythematosus. A review of the literature and presentation of 12 additional sets.* Am J Med, 1975. **59**(4): p. 533-52.
10. Deapen, D., et al., *A revised estimate of twin concordance in systemic lupus erythematosus.* Arthritis Rheum, 1992. **35**(3): p. 311-8.
11. Kuo, C.F., et al., *Familial Aggregation of Systemic Lupus Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families.* JAMA Intern Med, 2015. **175**(9): p. 1518-26.
12. Moser, K.L., et al., *Recent insights into the genetic basis of systemic lupus erythematosus.* Genes and Immunity, 2009. **10**(5): p. 373-379.
13. Kang, I., et al., *Defective control of latent Epstein-Barr virus infection in systemic lupus erythematosus.* Journal of Immunology, 2004. **172**(2): p. 1287-1294.
14. Bijl, M. and C.G.M. Kallenberg, *Ultraviolet light and cutaneous lupus.* Lupus, 2006. **15**(11): p. 724-727.
15. Gorelik, G., et al., *Impaired T Cell Protein Kinase Cδ Activation Decreases ERK Pathway Signaling in Idiopathic and Hydralazine-Induced Lupus.* The Journal of Immunology, 2007. **179**(8): p. 5553-5563.
16. Du, J., et al., *DNA methylation pathways and their crosstalk with histone methylation.* Nat Rev Mol Cell Biol, 2015. **16**(9): p. 519-32.
17. Costenbader, K.H., et al., *Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis.* Arthritis Rheum, 2004. **50**(3): p. 849-57.
18. Finckh, A., et al., *Occupational silica and solvent exposures and risk of systemic lupus erythematosus in urban women.* Arthritis Rheum, 2006. **54**(11): p. 3648-54.
19. Costenbader, K.H., et al., *Reproductive and menopausal factors and risk of systemic lupus erythematosus in women.* Arthritis Rheum, 2007. **56**(4): p. 1251-62.
20. Scofield, R.H., et al., *Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome.* Arthritis Rheum, 2008. **58**(8): p. 2511-7.
21. Crow, M.K., M. Olferiev, and K.A. Kirou, *Targeting of type I interferon in systemic autoimmune diseases.* Transl Res, 2015. **165**(2): p. 296-305.
22. Ettinger, R., et al., *IL-21 and BAFF/BLyS Synergize in Stimulating Plasma Cell Differentiation from a Unique Population of Human Splenic Memory B Cells.* The Journal of Immunology, 2007. **178**(5): p. 2872-2882.
23. Simpson, N., et al., *Expansion of circulating T cells resembling follicular helper T cells is a fixed phenotype that identifies a subset of severe systemic lupus erythematosus.* Arthritis Rheum, 2010. **62**(1): p. 234-44.
24. McKinney, E.F., et al., *A CD8+ T cell transcription signature predicts prognosis in autoimmune disease.* Nat Med, 2010. **16**(5): p. 586-91, 1p following 591.
25. Ohl, K. and K. Tenbrock, *Regulatory T cells in systemic lupus erythematosus.* European Journal of Immunology, 2015. **45**(2): p. 344-355.

26. Xing, Q., et al., *Elevated Th17 cells are accompanied by FoxP3+ Treg cells decrease in patients with lupus nephritis.* Rheumatol Int, 2012. **32**(4): p. 949-58.
27. Valencia, X., et al., *Deficient CD4+CD25high T regulatory cell function in patients with active systemic lupus erythematosus.* J Immunol, 2007. **178**(4): p. 2579-88.
28. Lyssuk, E.Y., et al., *Reduced number and function of CD4+CD25highFoxP3+ regulatory T cells in patients with systemic lupus erythematosus.* Adv Exp Med Biol, 2007. **601**: p. 113-9.
29. Bonelli, M., et al., *Quantitative and qualitative deficiencies of regulatory T cells in patients with systemic lupus erythematosus (SLE).* Int Immunol, 2008. **20**(7): p. 861-8.
30. Miyara, M., et al., *Global natural regulatory T cell depletion in active systemic lupus erythematosus.* J Immunol, 2005. **175**(12): p. 8392-400.
31. Habibagahi, M., et al., *Quantification of regulatory T cells in peripheral blood of patients with systemic lupus erythematosus.* Rheumatology International, 2011. **31**(9): p. 1219-1225.
32. Schmidt, A., et al., *Analysis of FOXP3(+) regulatory T cell subpopulations in peripheral blood and tissue of patients with systemic lupus erythematosus.* Immunologic Research, 2017. **65**(2): p. 551-563.
33. Alvarado-Sanchez, B., et al., *Regulatory T cells in patients with systemic lupus erythematosus.* Journal of Autoimmunity, 2006. **27**(2): p. 110-118.
34. Hiepe, F. and A. Radbruch, *Plasma cells as an innovative target in autoimmune disease with renal manifestations.* Nat Rev Nephrol, 2016. **12**(4): p. 232-40.
35. Jacobi, A.M., et al., *HLA-DRhigh/CD27high plasmablasts indicate active disease in patients with systemic lupus erythematosus.* Ann Rheum Dis, 2010. **69**(1): p. 305-8.
36. Munoz, L.E., et al., *The role of defective clearance of apoptotic cells in systemic autoimmunity.* Nat Rev Rheumatol, 2010. **6**(5): p. 280-9.
37. Lawrence, J.S., C.L. Martins, and G.L. Drake, *A family survey of lupus erythematosus. 1. Heritability.* J Rheumatol, 1987. **14**(5): p. 913-21.
38. Hom, G., et al., *Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX.* N Engl J Med, 2008. **358**(9): p. 900-9.
39. Kozyrev, S.V., et al., *Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus.* Nat Genet, 2008. **40**(2): p. 211-6.
40. Graham, R.R., et al., *Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus.* Nat Genet, 2008. **40**(9): p. 1059-61.
41. International Consortium for Systemic Lupus Erythematosus, G., et al., *Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.* Nat Genet, 2008. **40**(2): p. 204-10.
42. Morris, D.L., et al., *Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans.* Am J Hum Genet, 2012. **91**(5): p. 778-93.
43. Bentham, J., et al., *Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus.* Nat Genet, 2015. **47**(12): p. 1457-1464.
44. Demirci, F.Y., et al., *Identification of a New Susceptibility Locus for Systemic Lupus Erythematosus on Chromosome 12 in Individuals of European Ancestry.* Arthritis Rheumatol, 2016. **68**(1): p. 174-83.
45. Han, J.W., et al., *Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus.* Nat Genet, 2009. **41**(11): p. 1234-7.
46. Yang, W., et al., *Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus.* PLoS Genet, 2010. **6**(2): p. e1000841.
47. Okada, Y., et al., *A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese.* PLoS Genet, 2012. **8**(1): p. e1002455.
48. Lee, H.S., et al., *Ethnic specificity of lupus-associated loci identified in a genome-wide association study in Korean women.* Ann Rheum Dis, 2014. **73**(6): p. 1240-5.
49. Lessard, C.J., et al., *Identification of a Systemic Lupus Erythematosus Risk Locus Spanning ATG16L2, FCHSD2, and P2RY2 in Koreans.* Arthritis Rheumatol, 2016. **68**(5): p. 1197-1209.
50. Sun, C., et al., *High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry.* Nat Genet, 2016. **48**(3): p. 323-30.

51. Alarcon-Riquelme, M.E., et al., *Genome-Wide Association Study in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus Risk Loci and the Role of European Admixture.* Arthritis Rheumatol, 2016. **68**(4): p. 932-43.
52. Kaiser, R., et al., *Brief Report: Single-nucleotide polymorphisms in VKORC1 are risk factors for systemic lupus erythematosus in Asians.* Arthritis Rheum, 2013. **65**(1): p. 211-5.
53. Zhang, J., et al., *Gene-based meta-analysis of genome-wide association study data identifies independent single-nucleotide polymorphisms in ANXA6 as being associated with systemic lupus erythematosus in Asian populations.* Arthritis Rheumatol, 2015. **67**(11): p. 2966-77.
54. Morris, D.L., et al., *Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus.* Nat Genet, 2016. **48**(8): p. 940-946.
55. Molineros, J.E., et al., *Confirmation of five novel susceptibility loci for systemic lupus erythematosus (SLE) and integrated network analysis of 82 SLE susceptibility loci.* Hum Mol Genet, 2017. **26**(6): p. 1205-1216.
56. So, H.C., et al., *Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases.* Genet Epidemiol, 2011. **35**(5): p. 310-7.
57. Cortes, A. and M.A. Brown, *Promise and pitfalls of the Immunochip.* Arthritis Research & Therapy, 2011. **13**(1).
58. Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* Am J Hum Genet, 2011. **88**(1): p. 76-82.
59. Langefeld, C.D., et al., *Transancestral mapping and genetic load in systemic lupus erythematosus.* Nat Commun, 2017. **8**: p. 16021.
60. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
61. Wellcome Trust Case Control, C., et al., *Bayesian refinement of association signals for 14 loci in 3 common diseases.* Nat Genet, 2012. **44**(12): p. 1294-301.
62. Dixon, A.L., et al., *A genome-wide association study of global gene expression.* Nat Genet, 2007. **39**(10): p. 1202-7.
63. Choy, E., et al., *Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines.* PLoS Genet, 2008. **4**(11): p. e1000287.
64. Chen, L., D.L. Morris, and T.J. Vyse, *Genetic advances in systemic lupus erythematosus: an update.* Curr Opin Rheumatol, 2017.
65. Jansen, R.C. and J.P. Nap, *Genetical genomics: the added value from segregation.* Trends Genet, 2001. **17**(7): p. 388-91.
66. Brem, R.B., et al., *Genetic dissection of transcriptional regulation in budding yeast.* Science, 2002. **296**(5568): p. 752-5.
67. Cheung, V.G., et al., *Polymorphic cis- and trans-regulation of human gene expression.* PLoS Biol, 2010. **8**(9).
68. Morley, M., et al., *Genetic analysis of genome-wide variation in human gene expression.* Nature, 2004. **430**(7001): p. 743-7.
69. Nica, A.C. and E.T. Dermitzakis, *Expression quantitative trait loci: present and future.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1620): p. 20120362.
70. Nicolae, D.L., et al., *Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS.* Plos Genetics, 2010. **6**(4).
71. Moffatt, M.F., et al., *Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma.* Nature, 2007. **448**(7152): p. 470-U5.
72. Li, X., et al., *eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes.* Allergy, 2015. **70**(10): p. 1309-18.
73. Fairfax, B.P., et al., *Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles.* Nat Genet, 2012. **44**(5): p. 502-10.
74. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
75. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations.* Nature Genetics, 2013. **45**(10): p. 1238-U195.
76. Consortium, G.T., et al., *Genetic effects on gene expression across human tissues.* Nature, 2017. **550**(7675): p. 204-213.
77. Nica, A.C., et al., *Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations.* PLoS Genet, 2010. **6**(4): p. e1000895.

78. Giambartolomei, C., et al., *Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics.* Plos Genetics, 2014. **10**(5).
79. Rockman, M.V. and L. Kruglyak, *Genetics of global gene expression.* Nat Rev Genet, 2006. **7**(11): p. 862-72.
80. Farh, K.K., et al., *Genetic and epigenetic fine mapping of causal autoimmune disease variants.* Nature, 2015. **518**(7539): p. 337-43.
81. Albert, F.W. and L. Kruglyak, *The role of regulatory variation in complex traits and disease.* Nat Rev Genet, 2015. **16**(4): p. 197-212.
82. Cheung, V.G., et al., *Natural variation in human gene expression assessed in lymphoblastoid cells.* Nat Genet, 2003. **33**(3): p. 422-5.
83. Monks, S.A., et al., *Genetic inheritance of gene expression in human cell lines.* Am J Hum Genet, 2004. **75**(6): p. 1094-105.
84. Plagnol, V., et al., *Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses.* PLoS One, 2008. **3**(8): p. e2966.
85. Zhang, X., et al., *Identification of common genetic variants controlling transcript isoform variation in human whole blood.* Nat Genet, 2015. **47**(4): p. 345-52.
86. Joehanes, R., et al., *Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies.* Genome Biol, 2017. **18**(1): p. 16.
87. Fairfax, B.P. and J.C. Knight, *Genetics of gene expression in immunity to infection.* Curr Opin Immunol, 2014. **30**: p. 63-71.
88. Ishigaki, K., et al., *Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis.* Nat Genet, 2017. **49**(7): p. 1120-1125.
89. Westra, H.J., et al., *Cell Specific eQTL Analysis without Sorting Cells.* PLoS Genet, 2015. **11**(5): p. e1005223.
90. Zhernakova, D.V., et al., *Identification of context-dependent expression quantitative trait loci in whole blood.* Nat Genet, 2017. **49**(1): p. 139-145.
91. Fairfax, B.P., et al., *Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression.* Science, 2014. **343**(6175): p. 1246949.
92. Zeller, T., et al., *Genetics and beyond--the transcriptome of human monocytes and disease susceptibility.* PLoS One, 2010. **5**(5): p. e10693.
93. Sarda, S. and S. Hannenhalli, *Next-generation sequencing and epigenomics research: a hammer in search of nails.* Genomics Inform, 2014. **12**(1): p. 2-11.
94. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
95. Belton, J.M., et al., *Hi-C: a comprehensive technique to capture the conformation of genomes.* Methods, 2012. **58**(3): p. 268-76.
96. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.
97. Astle, W.J., et al., *The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease.* Cell, 2016. **167**(5): p. 1415-1429 e19.
98. Javierre, B.M., et al., *Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters.* Cell, 2016. **167**(5): p. 1369-1384 e19.
99. Chen, L., et al., *Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells.* Cell, 2016. **167**(5): p. 1398-1414 e24.
100. Maecker, H.T., J.P. McCoy, and R. Nussenblatt, *Standardizing immunophenotyping for the Human Immunology Project.* Nat Rev Immunol, 2012. **12**(3): p. 191-200.
101. Spitz, F. and E.E. Furlong, *Transcription factors: from enhancer binding to developmental control.* Nat Rev Genet, 2012. **13**(9): p. 613-26.
102. Rebollo, A. and C. Schmitt, *Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies.* Immunology and cell biology, 2003. **81**(3): p. 171-175 %@ 0818-9641.
103. Yap, W.H., et al., *STAT4 is a target of the hematopoietic zinc-finger transcription factor Ikaros in T cells.* Febs Letters, 2005. **579**(20): p. 4470-4478.
104. Ghadiri, A., et al., *Critical function of Ikaros in controlling Aiolos gene expression.* FEBS letters, 2007. **581**(8): p. 1605-1616 %@ 0014-5793.
105. Aad, G., et al., *Measurement of dijet azimuthal decorrelations in pp collisions at sqrt(s)=7 TeV.* Phys Rev Lett, 2011. **106**(17): p. 172002.
106. Joshi, I., et al., *Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia.* Nat Immunol, 2014. **15**(3): p. 294-304.

107. Medeiros, B.C., *Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia.* N Engl J Med, 2009. **360**(17): p. 1787; author reply 1787-8.
108. Cunninghame Graham, D.S., et al., *Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus.* PLoS Genet, 2011. **7**(10): p. e1002341.
109. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nature genetics, 2010. **42**(12): p. 1118-1125 %@ 1061-4036.
110. Ueta, M., et al., *IKZF1, a new susceptibility gene for cold medicine-related Stevens-Johnson syndrome/toxic epidermal necrolysis with severe mucosal involvement.* J Allergy Clin Immunol, 2015. **135**(6): p. 1538-45 e17.
111. Swafford, A.D.E., et al., *An allele of IKZF1 (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes.* Diabetes, 2011. **60**(3): p. 1041-1044 %@ 0012-1797.
112. Thornton, A.M., et al., *Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory cells.* J Immunol, 2010. **184**(7): p. 3433-41.
113. Getnet, D., et al., *A role for the transcription factor Helios in human CD4(+)CD25(+) regulatory T cells.* Mol Immunol, 2010. **47**(7-8): p. 1595-600.
114. Alexander, T., et al., *Foxp3+ Helios+ regulatory T cells are expanded in active systemic lupus erythematosus.* Ann Rheum Dis, 2013. **72**(9): p. 1549-58.
115. Morgan, B., et al., *Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation.* Embo Journal, 1997. **16**(8): p. 2004-2013.
116. Romero, F., et al., *Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization.* The EMBO journal, 1999. **18**(12): p. 3419-3430 %@ 0261-4189.
117. Sun, J., et al., *Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice.* Journal of Immunology, 2003. **170**(4): p. 1699-1706.
118. Marinho, S., et al., *17q12-21 Variants are associated with asthma and interact with active smoking in an adult population from the United Kingdom.* Annals of Allergy Asthma & Immunology, 2012. **108**(6): p. 402-+.
119. Kurreeman, F.A.S., et al., *Use of a Multiethnic Approach to Identify Rheumatoid-Arthritis-Susceptibility Loci, 1p36 and 17q12.* American Journal of Human Genetics, 2012. **90**(3): p. 524-532.
120. Qiu, R., et al., *Genetic variants on 17q21 are associated with ankylosing spondylitis susceptibility and severity in a Chinese Han population.* Scandinavian Journal of Rheumatology, 2013. **42**(6): p. 469-472.
121. Croft, M., *Control of immunity by the TNFR-related molecule OX40 (CD134).* Annu Rev Immunol, 2010. **28**: p. 57-78.
122. Murata, K., et al., *Impairment of antigen-presenting cell function in mice lacking expression of OX40 ligand.* The Journal of experimental medicine, 2000. **191**(2): p. 365-374 %@ 0022-1007.
123. Ohshima, Y., et al., *Expression and function of OX40 ligand on human dendritic cells.* The Journal of Immunology, 1997. **159**(8): p. 3838-3848 %@ 0022-1767.
124. Sato, T., et al., *Consequences of OX40-OX40 ligand interactions in Langerhans cell function: enhanced contact hypersensitivity responses in OX40L-transgenic mice.* European journal of immunology, 2002. **32**(11): p. 3326-3335 %@ 1521-4141.
125. Weinberg, A.D., et al., *Blocking OX-40/OX-40 ligand interaction in vitro and in vivo leads to decreased T cell function and amelioration of experimental allergic encephalomyelitis.* The Journal of Immunology, 1999. **162**(3): p. 1818-1826 %@ 0022-1767.
126. Song, J., et al., *The costimulation-regulated duration of PKB activation controls T cell longevity.* Nature immunology, 2004. **5**(2): p. 150-158.
127. Croft, M., et al., *The significance of OX40 and OX40L to T-cell biology and immune disease.* Immunological Reviews, 2009. **229**: p. 173-191.
128. Graham, D.S.C., et al., *Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus.* Nature Genetics, 2008. **40**(1): p. 83-89.
129. Zhang, X., et al., *Identification of common genetic variants controlling transcript isoform variation in human whole blood.* Nature genetics, 2015. **47**(4): p. 345-352 %@ 1061-4036.
130. Ito, T., et al., *OX40 ligand shuts down IL-10-producing regulatory T cells.* Proceedings of the National Academy of Sciences, 2006. **103**(35): p. 13138-13143 %@ 0027-8424.

131. Davidson, A., *What is damaging the kidney in lupus nephritis?* Nat Rev Rheumatol, 2016. **12**(3): p. 143-53.
132. Peschken, C.A., et al., *The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort.* J Rheumatol, 2009. **36**(6): p. 1200-8.
133. Pons-Estel, B.A., et al., *The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics".* Medicine (Baltimore), 2004. **83**(1): p. 1-17.
134. Alarcon, G.S., et al., *Baseline characteristics of a multiethnic lupus cohort: PROFILE.* Lupus, 2002. **11**(2): p. 95-101.
135. Belsky, D.W., et al., *Development and evaluation of a genetic risk score for obesity.* Biodemography Soc Biol, 2013. **59**(1): p. 85-100.
136. De Jager, P.L., et al., *Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score.* The Lancet Neurology, 2009. **8**(12): p. 1111-1119.
137. Karlson, E.W., et al., *Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk.* Ann Rheum Dis, 2010. **69**(6): p. 1077-85.
138. Yarwood, A., et al., *A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk.* Ann Rheum Dis, 2015. **74**(1): p. 170-6.
139. Dooley, M.A., C. Aranow, and E.M. Ginzler, *Review of ACR renal criteria in systemic lupus erythematosus.* Lupus, 2004. **13**(11): p. 857-60.
140. O'Connell, J., et al., *Haplotype estimation for biobank-scale data sets.* Nat Genet, 2016. **48**(7): p. 817-20.
141. Marchini, J., et al., *A new multipoint method for genome-wide association studies by imputation of genotypes.* Nat Genet, 2007. **39**(7): p. 906-13.
142. Yang, H. and K. Wang, *Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR.* Nat Protoc, 2015. **10**(10): p. 1556-66.
143. Iotchkova, V., et al., *Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps.* Nat Genet, 2016. **48**(11): p. 1303-1312.
144. Iotchkova, V., et al., *GARFIELD - GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction.* bioRxiv, 2016.
145. Consortium, U.K., et al., *The UK10K project identifies rare variants in health and disease.* Nature, 2015. **526**(7571): p. 82-90.
146. Trynka, G., et al., *Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci.* Am J Hum Genet, 2015. **97**(1): p. 139-52.
147. Raj, T., et al., *Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes.* Science, 2014. **344**(6183): p. 519-23.
148. Lloyd-Jones, L.R., et al., *The Genetic Architecture of Gene Expression in Peripheral Blood.* Am J Hum Genet, 2017. **100**(2): p. 228-237.
149. Stranger, B.E., et al., *Patterns of cis regulatory variation in diverse human populations.* PLoS Genet, 2012. **8**(4): p. e1002639.
150. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations.* Bioinformatics, 2012. **28**(10): p. 1353-8.
151. Zuk, O., et al., *The mystery of missing heritability: Genetic interactions create phantom heritability.* Proc Natl Acad Sci U S A, 2012. **109**(4): p. 1193-8.
152. Kidd, J.M., et al., *Mapping and sequencing of structural variation from eight human genomes.* Nature, 2008. **453**(7191): p. 56-64.
153. Price, A.L., C.C. Spencer, and P. Donnelly, *Progress and promise in understanding the genetic basis of common diseases.* Proc Biol Sci, 2015. **282**(1821): p. 20151684.
154. Myers, S., et al., *A fine-scale map of recombination rates and hotspots across the human genome.* Science, 2005. **310**(5746): p. 321-4.
155. Hughes, T., et al., *Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus.* Ann Rheum Dis, 2012. **71**(5): p. 694-9.
156. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.* Biometrics, 1988. **44**(3): p. 837-45.
157. Bombardier, C., et al., *Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE.* Arthritis Rheum, 1992. **35**(6): p. 630-40.

158. Benjamini, Y., et al., *Controlling the false discovery rate in behavior genetics research.* Behav Brain Res, 2001. **125**(1-2): p. 279-84.
159. Dubois, B., et al., *Revising the definition of Alzheimer's disease: a new lexicon.* Lancet Neurol, 2010. **9**(11): p. 1118-27.
160. Yang, F., et al., *Identifying cis-mediators for trans-eQTLs across many human tissues using genomic mediation analysis.* Genome Research, 2017. **27**(11): p. 1859-1871.
161. Fehrmann, R.S.N., et al., *Trans-eQTLs Reveal That Independent Genetic Variants Associated with a Complex Phenotype Converge on Intermediate Genes, with a Major Role for the HLA.* Plos Genetics, 2011. **7**(8).
162. Goring, H.H.H., et al., *Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes.* Nature Genetics, 2007. **39**(10): p. 1208-1216.
163. Stranger, B.E., et al., *Relative impact of nucleotide and copy number variation on gene expression phenotypes.* Science, 2007. **315**(5813): p. 848-853.
164. Fu, J., et al., *System-wide molecular evidence for phenotypic buffering in Arabidopsis.* Nature Genetics, 2009. **41**(2): p. 166-167.
165. Strange, A., et al., *A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1.* Nature Genetics, 2010. **42**(11): p. 985-U106.
166. Rajsbaum, R., et al., *Unanchored K48-linked polyubiquitin synthesized by the E3-ubiquitin ligase TRIM6 stimulates the interferon-IKKepsilon kinase-mediated antiviral response.* Immunity, 2014. **40**(6): p. 880-95.
167. Rhodes, B., et al., *The rs1143679 (R77H) lupus associated variant of ITGAM (CD11b) impairs complement receptor 3 mediated functions in human monocytes.* Ann Rheum Dis, 2012. **71**(12): p. 2028-34.
168. Anaya, J.M., et al., *Evaluation of genetic association between an ITGAM non-synonymous SNP (rs1143679) and multiple autoimmune diseases.* Autoimmunity Reviews, 2012. **11**(4): p. 276-280.
169. Ravetch, J.V. and S. Bolland, *IgG Fc receptors.* Annual Review of Immunology, 2001. **19**: p. 275-290.
170. Li, X.R. and R.P. Kimberly, *Targeting the Fc receptor in autoimmune disease.* Expert Opinion on Therapeutic Targets, 2014. **18**(3): p. 335-350.
171. Li, Y.R. and B.J. Keating, *Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations.* Genome Medicine, 2014. **6**.
172. Magi, R., et al., *Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution.* Human Molecular Genetics, 2017. **26**(18): p. 3639-3650.
173. Campbell, M.C. and S.A. Tishkoff, *African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping.* Annual Review of Genomics and Human Genetics, 2008. **9**: p. 403-433.
174. Gunther, C., et al., *Defective removal of ribonucleotides from DNA promotes systemic autoimmunity.* Journal of Clinical Investigation, 2015. **125**(1): p. 413-424.
175. Bendix, I., et al., *MAPK3 deficiency drives autoimmunity via DC arming.* European Journal of Immunology, 2010. **40**(5): p. 1486-1495.
176. Coustet, B., et al., *Association study of ITGAM, ITGAX, and CD58 autoimmune risk loci in systemic sclerosis: results from 2 large European Caucasian cohorts.* J Rheumatol, 2011. **38**(6): p. 1033-8.
177. Qi, Y.M., et al., *IFI6 Inhibits Apoptosis via Mitochondrial-Dependent Pathway in Dengue Virus 2 Infected Vascular Endothelial Cells (vol 10, e0132743, 2015).* Plos One, 2015. **10**(9).
178. Green, D.S., H.A. Young, and J.C. Valencia, *Current prospects of type II interferon gamma signaling and autoimmunity.* J Biol Chem, 2017. **292**(34): p. 13925-13933.
179. Zoledziewska, M., et al., *Variation within the CLEC16A gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia.* Genes and Immunity, 2009. **10**(1): p. 15-17.
180. Grundberg, E., et al., *Mapping cis- and trans-regulatory effects across multiple tissues in twins.* Nat Genet, 2012. **44**(10): p. 1084-9.
181. Lappalainen, T., et al., *Transcriptome and genome sequencing uncovers functional variation in humans.* Nature, 2013. **501**(7468): p. 506-511.
182. Small, K.S., et al., *Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes (vol 43, pg 561, 2011).* Nature Genetics, 2011. **43**(10): p. 1040-1040.

183. Lewis, M.J., et al., *UBE2L3 polymorphism amplifies NF-kappaB activation and promotes plasma cell development, linking linear ubiquitination to multiple autoimmune diseases.* Am J Hum Genet, 2015. **96**(2): p. 221-34.
184. Odhams, C.A., et al., *Mapping eQTLs with RNA-seq reveals novel susceptibility genes, non-coding RNAs and alternative-splicing events in systemic lupus erythematosus.* Hum Mol Genet, 2017. **26**(5): p. 1003-1017.
185. Odhams, C.A., D.S.C. Graham, and T.J. Vyse, *Profiling RNA-Seq at multiple resolutions markedly increases the number of causal eQTLs in autoimmune disease.* Plos Genetics, 2017. **13**(10).
186. Peng, S.L., *Experimental use of mouse models of systemic lupus erythematosus*, in *Autoimmunity*. 2012, Springer. p. 135-168.
187. Du, Y., et al., *Animal Models of Lupus and Lupus Nephritis.* Current Pharmaceutical Design, 2015. **21**(18): p. 2320-2349.
188. Flam, F., *Hints of a language in junk DNA.* Science, 1994. **266**(5189): p. 1320.
189. Pennisi, E., *Genomics. ENCODE project writes eulogy for junk DNA.* Science, 2012. **337**(6099): p. 1159, 1161.
190. Gerstein, M.B., et al., *Architecture of the human regulatory network derived from ENCODE data.* Nature, 2012. **489**(7414): p. 91-100.
191. Boyle, A.P., et al., *High-resolution mapping and characterization of open chromatin across the genome.* Cell, 2008. **132**(2): p. 311-22.
192. Liu, Z., Y. Zou, and A. Davidson, *Plasma cells in systemic lupus erythematosus: the long and short of it all.* Eur J Immunol, 2011. **41**(3): p. 588-91.
193. Wu, H. and B.D. Humphreys, *The promise of single-cell RNA sequencing for kidney disease investigation.* Kidney Int, 2017. **92**(6): p. 1334-1342.
194. Eberwine, J., et al., *The promise of single-cell sequencing.* Nature Methods, 2014. **11**(1): p. 25-27.
195. Cervera, R., et al., *Morbidity and mortality in systemic lupus erythematosus during a 10-year period - A comparison of early and late manifestations in a cohort of 1,000 patients.* Medicine, 2003. **82**(5): p. 299-308.
196. Font, J., et al., *Clusters of clinical and immunologic features in systemic lupus erythematosus: Analysis of 600 patients from a single center.* Seminars in Arthritis and Rheumatism, 2004. **33**(4): p. 217-230.
197. Bello, G.A., et al., *Development and validation of a simple lupus severity index using ACR criteria for classification of SLE.* Lupus Sci Med, 2016. **3**(1): p. e000136.
198. Lazaro, D., *Elderly-onset systemic lupus erythematosus: prevalence, clinical course and treatment.* Drugs Aging, 2007. **24**(9): p. 701-15.
199. Janwityanujit, S., et al., *Age-related differences on clinical and immunological manifestations of SLE.* Asian Pac J Allergy Immunol, 1995. **13**(2): p. 145-9.
200. Tomic-Lucic, A., et al., *Late-onset systemic lupus erythematosus: clinical features, course, and prognosis.* Clin Rheumatol, 2013. **32**(7): p. 1053-8.
201. Cortini, A., et al., *B cell OX40L supports T follicular helper cell development and contributes to SLE pathogenesis.* Ann Rheum Dis, 2017. **76**(12): p. 2095-2103.
202. Hoshino, A., et al., *Abnormal hematopoiesis and autoimmunity in human subjects with germline IKZF1 mutations.* J Allergy Clin Immunol, 2017. **140**(1): p. 223-231.
203. Tajuddin, S.M., et al., *Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases.* Am J Hum Genet, 2016. **99**(1): p. 22-39.
204. Zhong, H. and K. Yazdanbakhsh, *Differential control of Helios(+/-) Treg development by monocyte subsets through disparate inflammatory cytokines.* Blood, 2013. **121**(13): p. 2494-2502.
205. Nakayama, Y., et al., *Aiolos Overexpression in Systemic Lupus Erythematosus B Cell Subtypes and BAFF-Induced Memory B Cell Differentiation Are Reduced by CC-220 Modulation of Cereblon Activity.* Journal of Immunology, 2017. **199**(7): p. 2388-2407.
206. Finak, G., et al., *Standardizing Flow Cytometry Immunophenotyping Analysis from the Human ImmunoPhenotyping Consortium.* Scientific Reports, 2016. **6**.
207. Jacquemin, C., et al., *OX40 Ligand Contributes to Human Lupus Pathogenesis by Promoting T Follicular Helper Response.* Immunity, 2015. **42**(6): p. 1159-1170.
208. Okada, Y., et al., *Genetics of rheumatoid arthritis contributes to biology and drug discovery.* Nature, 2014. **506**(7488): p. 376-+.
209. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nature genetics, 2010. **42**(12): p. 1118.

210. Jostins, L., et al., *Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119.
211. Liu, J.Z., et al., *Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations.* Nature genetics, 2015. **47**(9): p. 979.
212. Beecham, A.H., et al., *Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis.* Nature genetics, 2013. **45**(11): p. 1353.
213. Cordell, H.J., et al., *International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways.* Nature communications, 2015. **6**: p. 8019.
214. Stahl, E.A., et al., *Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci.* Nature genetics, 2010. **42**(6): p. 508.
215. Eyre, S., et al., *High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis.* Nature genetics, 2012. **44**(12): p. 1336.
216. Onengut-Gumuscu, S., et al., *Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers.* Nature genetics, 2015. **47**(4): p. 381.
217. Anderson, C.A., et al., *Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47.* Nature genetics, 2011. **43**(3): p. 246.
218. Pai, A.A., et al., *The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels.* PLoS genetics, 2012. **8**(10): p. e1003000.
219. Hu, W., et al., *Down-regulated expression of IKZF1 mRNA in peripheral blood mononuclear cells from patients with systemic lupus erythematosus.* Rheumatol Int, 2011. **31**(6): p. 819-22.
220. Sellars, M., P. Kastner, and S. Chan, *Ikaros in B cell development and function.* World J Biol Chem, 2011. **2**(6): p. 132-9.
221. Sridharan, R. and S.T. Smale, *Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes.* Journal of Biological Chemistry, 2007. **282**(41): p. 30227-30238.
222. Koipally, J. and K. Georgopoulos, *Ikaros interactions with CtBP reveal a repression mechanism that is independent of histone deacetylase activity.* Journal of Biological Chemistry, 2000. **275**(26): p. 19594-19602.
223. Koipally, J. and K. Georgopoulos, *Ikaros-CtIP interactions do not require C-terminal binding protein and participate in a deacetylase-independent mode of repression.* Journal of Biological Chemistry, 2002. **277**(26): p. 23143-23149.
224. O'Neill, D.W., et al., *An ikaros-containing chromatin-remodeling complex in adult-type erythroid cells.* Molecular and cellular biology, 2000. **20**(20): p. 7572-7582.
225. Wang, W., *The SWI/SNF family of ATP-dependent chromatin remodelers: similar mechanisms for diverse functions*, in *Protein Complexes that Modify Chromatin*. 2003, Springer. p. 143-169.
226. Wang, J.-H., et al., *Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation.* Immunity, 1996. **5**(6): p. 537-549.
227. Yoshida, T., et al., *Early hematopoietic lineage restrictions directed by Ikaros.* Nature immunology, 2006. **7**(4): p. 382.
228. Reynaud, D., et al., *Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros.* Nature immunology, 2008. **9**(8): p. 927.
229. Kirstetter, P., et al., *Ikaros is critical for B cell differentiation and function.* European journal of immunology, 2002. **32**(3): p. 720-730.
230. Kim, H.-J., et al., *Stable inhibitory activity of regulatory T cells requires the transcription factor Helios.* Science, 2015. **350**(6258): p. 334-339.
231. Getnet, D., et al., *A role for the transcription factor Helios in human CD4+ CD25+ regulatory T cells.* Molecular immunology, 2010. **47**(7-8): p. 1595-1600.
232. Golding, A., et al., *The percentage of FoxP3+ Helios+ Treg cells correlates positively with disease activity in systemic lupus erythematosus.* Arthritis & Rheumatology, 2013. **65**(11): p. 2898-2906.
233. Karrar, S. and D.S. Cunninghame Graham, *Abnormal B-cell development in Systemic Lupus Erythematosus: what the genetics tell us.* Arthritis & Rheumatology, 2017.
234. Isnardi, I., et al., *IRAK-4-and MyD88-dependent pathways are essential for the removal of developing autoreactive B cells in humans.* Immunity, 2008. **29**(5): p. 746-757.

235. Cai, X., et al., *Overexpression of Aiolos in peripheral blood mononuclear cell subsets from patients with systemic lupus erythematosus and rheumatoid arthritis.* Biochemical genetics, 2016. **54**(1): p. 73-82.

Appendix A. Chapter 2 Supplementary Tables

Supplementary Table 2-1. 424 DHS hotspots from ENCODE and Roadmap for functional enrichment analysis in GARFIELD.

Tissue	Celltype	Annotation	Tissue	Celltype	Annotation
epithelium	A549	A549-all.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Left	UW.Fetal_Kidney_Left.ChromatinAccessibility.H-24089.DS18466.twopass.merge150.wgt10.zgt2.wig
blood	Adult_CD4+	Adult_ThO_AllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Right	UW.Fetal_Kidney_Right.ChromatinAccessibility.H-23435.DS15651.twopass.merge150.wgt10.zgt2.wig
skin	AG04449	AG04449-DS12319.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Right	UW.Fetal_Kidney_Right.ChromatinAccessibility.H-23589.DS16441.twopass.merge150.wgt10.zgt2.wig
lung	AG04450	AG04450-DS12270.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Right	UW.Fetal_Kidney_Right.ChromatinAccessibility.H-23640.DS16801.twopass.merge150.wgt10.zgt2.wig
skin	AG09309	AG09309-DS12352.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Right	UW.Fetal_Kidney_Right.ChromatinAccessibility.H-23758.DS17144.twopass.merge150.wgt10.zgt2.wig
gingival	AG09319	AG09319-DS12291.twopass.merge150.wgt10.zgt2.wig	fetal_kidney	Fetal_Kidney_Right	UW.Fetal_Kidney_Right.ChromatinAccessibility.H-24089.DS18463.twopass.merge150.wgt10.zgt2.wig
skin	AG10803	AG10803-DS12384.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-22676.DS12646.twopass.merge150.wgt10.zgt2.wig
blood_vessel	AoAF	AoAF-DS13523.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-22727.DS12817.twopass.merge150.wgt10.zgt2.wig
brain	BE2_C	BE_2_C-DS14625.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-22934B.DS13507.twopass.merge150.wgt10.zgt2.wig
skin	BJ	BJ-DS10081.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23090.DS13985.twopass.merge150.wgt10.zgt2.wig
colon	Caco-2	CACO2-DS8235.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23247.DS14666.twopass.merge150.wgt10.zgt2.wig
blood	CD14+	CD14-DS18065.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23266.DS14724.twopass.merge150.wgt10.zgt2.wig
blood	CD20+	CD20-DS17541.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23266.DS14751.twopass.merge150.wgt10.zgt2.wig
blood	CD34+	CD34-DS16814.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23284.DS14809.twopass.merge150.wgt10.zgt2.wig
blood	CMK	CMK-DS12393.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23284.DS14820.twopass.merge150.wgt10.zgt2.wig
muscle	E_myoblast	E_myoblast_AllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23365.DS15227.twopass.merge150.wgt10.zgt2.wig
blood	GM06990	GM06990-DS7748.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23399.DS15461.twopass.merge150.wgt10.zgt2.wig
blood	GM12864	GM12864-DS12431.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung	UW.Fetal_Lung.ChromatinAccessibility.H-23419.DS15573.twopass.merge150.wgt10.zgt2.wig
blood	GM12865	GM12865-DS12436.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23435.DS15637.twopass.merge150.wgt10.zgt2.wig
blood	GM12878	GM12878-all.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23604.DS16570.twopass.merge150.wgt10.zgt2.wig
cerebellar	HAc	HAc-DS14765.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23640.DS18170.twopass.merge150.wgt10.zgt2.wig
epithelium	HAEpiC	HAEpiC-DS12663.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23744.DS17105.twopass.merge150.wgt10.zgt2.wig
brain_hippocampus	HA-h	HAh-DS15192.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23758.DS17154.twopass.merge150.wgt10.zgt2.wig
spinal_cord	HA-sp	HAsp-DS14790.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23833.DS17444.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HBMEC	HBMEC-DS13817.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23887.DS17674.twopass.merge150.wgt10.zgt2.wig
heart	HCFaa	HCFaa-DS13480.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23914.DS17739.twopass.merge150.wgt10.zgt2.wig
heart	HCF	HCF-DS12501.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-23964.DS17835.twopass.merge150.wgt10.zgt2.wig
heart	HCM	HCM-DS12599.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-24005.DS17959.twopass.merge150.wgt10.zgt2.wig
eye	HConF	HConF-DS11642.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-24089.DS18421.twopass.merge150.wgt10.zgt2.wig
epithelium	HCPEpiC	HCPEpiC-DS12447.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Left	UW.Fetal_Lung_Left.ChromatinAccessibility.H-24111.DS18487.twopass.merge150.wgt10.zgt2.wig
colon	HCT-116	HCT116-DS13551.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23435.DS15632.twopass.merge150.wgt10.zgt2.wig
epithelium	HEEpiC	HEEpiC-DS12763.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23604.DS16566.twopass.merge150.wgt10.zgt2.wig
cervix	HeLa-S3	Hela-all.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23640.DS16790.twopass.merge150.wgt10.zgt2.wig
liver	HepG2	HepG2-all.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23744.DS17101.twopass.merge150.wgt10.zgt2.wig
es_cell	HESC	HESC-all.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23758.DS17162.twopass.merge150.wgt10.zgt2.wig
es_cell	hESCT0	hESCT0-DS11909.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23887.DS17670.twopass.merge150.wgt10.zgt2.wig
foreskin	HFF	HFF-DS15115.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-23964.DS17831.twopass.merge150.wgt10.zgt2.wig
foreskin	HFF-Myc	HFF_MyC-DS15079.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-24005.DS17954.twopass.merge150.wgt10.zgt2.wig
gingival	HGF	HGF-DS11752.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-24089.DS18418.twopass.merge150.wgt10.zgt2.wig
epithelium	HIPEpiC	HIPEpiC-DS12684.twopass.merge150.wgt10.zgt2.wig	fetal_lung	Fetal_Lung_Right	UW.Fetal_Lung_Right.ChromatinAccessibility.H-24111.DS18492.twopass.merge150.wgt10.zgt2.wig
blood	HL-60	HL60-DS11809.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-23808.DS17432.twopass.merge150.wgt10.zgt2.wig
breast	HMEC	HMEC-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-23914.DS17765.twopass.merge150.wgt10.zgt2.wig
breast	HMF	HMF-DS13363.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-23941.DS17825.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dAd	HMVEC_dAd-DS12957.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-23964.DS17848.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dBl-Ad	HMVEC_dBlAd-DS13337.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24005.DS18080.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dBl-Neo	HMVEC_dBlNeo-DS13242.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24042.DS18176.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dLy-Ad	HMVEC_dLyAd-DS13261.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24078.DS18379.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dLy-Neo	HMVEC_dLyNeo-DS13150.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24089.DS18452.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-dNeo	HMVEC_dNeo-DS12937.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24111.DS18473.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-LBl	HMVEC_LBl-DS13372.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24125.DS18559.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HMVEC-LLy	HMVEC_Lly-DS13185.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24143.DS18992.twopass.merge150.wgt10.zgt2.wig
epithelium	HNPCEpiC	HNPCEpiC-DS12467.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24198.DS19051.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HPAEC	HPAEC-DS12916.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24218.DS19053.twopass.merge150.wgt10.zgt2.wig

Tissue	Celltype	Annotation	Tissue	Celltype	Annotation
blood_vessel	HPAF	HPAF-DS13411.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24244.DS19270.twopass.merge150.wgt10.zgt2.wig
epithelium	HPdLF	HPdLF-DS13573.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24259.DS19295.twopass.merge150.wgt10.zgt2.wig
lung	HPF	HPF-DS13390.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24272.DS19380.twopass.merge150.wgt10.zgt2.wig
epithelium	HRCEpiC	HRCE-DS10666.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Arm	UW.Fetal_Muscle_Arm.ChromatinAccessibility.H-24297.DS19646.twopass.merge150.wgt10.zgt2.wig
epithelium	HRE	HRE-DS10641.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-23914.DS17767.twopass.merge150.wgt10.zgt2.wig
kidney	HRGEC	HRGEC-DS13716.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-23964.DS17850.twopass.merge150.wgt10.zgt2.wig
epithelium	HRPEpiC	HRPEpiC-DS12583.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24005.DS18083.twopass.merge150.wgt10.zgt2.wig
muscle	HSMM	HSMM-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24078.DS18377.twopass.merge150.wgt10.zgt2.wig
muscle	HSMM	HSMM_D-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24089.DS18454.twopass.merge150.wgt10.zgt2.wig
blood	Th1	hTH1-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24111.DS18468.twopass.merge150.wgt10.zgt2.wig
blood	Th2	hTH2-DS7842.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24143.DS18842.twopass.merge150.wgt10.zgt2.wig
blood_vessel	HUVEC	HUVEC-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24218.DS19117.twopass.merge150.wgt10.zgt2.wig
connective	HVMF	HVMF-DS13981.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24244.DS19283.twopass.merge150.wgt10.zgt2.wig
uterus	Ishikawa	Ishikawa_E_AllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24272.DS19384.twopass.merge150.wgt10.zgt2.wig
uterus	Ishikawa	Ishikawa_T_AllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24279.DS19441.twopass.merge150.wgt10.zgt2.wig
blood	Jurkat	Jurkat-DS12659.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24297.DS19648.twopass.merge150.wgt10.zgt2.wig
blood	K562	K562-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Back	UW.Fetal_Muscle_Back.ChromatinAccessibility.H-24409.DS20244.twopass.merge150.wgt10.zgt2.wig
prostate	LNCaP	LNCap-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-23808.DS17429.twopass.merge150.wgt10.zgt2.wig
breast	MCF-7	MCF7-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24078.DS18386.twopass.merge150.wgt10.zgt2.wig
blood	NB4	NB4-DS12543.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24089.DS18456.twopass.merge150.wgt10.zgt2.wig
nervous	NH-A	NHA-DS12800.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24111.DS18471.twopass.merge150.wgt10.zgt2.wig
skin	NHDF-Ad	NHDF_Ad-DS12863.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24143.DS18844.twopass.merge150.wgt10.zgt2.wig
skin	NHDF-neo	NHDF_Neo-DS11923.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24198.DS19158.twopass.merge150.wgt10.zgt2.wig
skin	NHEK	NHEK-all.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24218.DS19115.twopass.merge150.wgt10.zgt2.wig
lung	NHLF	NHLF-DS12829.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24244.DS19272.twopass.merge150.wgt10.zgt2.wig
testis	NT2-D1	NT2_D1-DS14575.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24259.DS19291.twopass.merge150.wgt10.zgt2.wig
pancreas	PANC-1	PANC1-DS9955.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24279.DS19436.twopass.merge150.wgt10.zgt2.wig
prostate	PrEC	PrEC-DS12098.hg19.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24297.DS19643.twopass.merge150.wgt10.zgt2.wig
epithelium	RPTEC	RPTEC-DS14061.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Leg	UW.Fetal_Muscle_Leg.ChromatinAccessibility.H-24409.DS20239.twopass.merge150.wgt10.zgt2.wig
prostate	RWPE1	RWPE_AllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_muscle_lower_limb	Fetal_Muscle_Lower_Limb_Skeletal	UW.Fetal_Muscle_Lower_Limb_Skeletal.ChromatinAccessibility.H-24042.DS18174.twopass.merge150.wgt10.zgt2.wig
epithelium	SAEC	SAEC-DS10518.twopass.merge150.wgt10.zgt2.wig	fetal_muscle_trunk	Fetal_Muscle_Trunk	UW.Fetal_Muscle_Trunk.ChromatinAccessibility.H-23941.DS17827.twopass.merge150.wgt10.zgt2.wig
muscle	SKMC	SkMC-DS11949.twopass.merge150.wgt10.zgt2.wig	fetal_muscle_trunk	Fetal_Muscle_Trunk	UW.Fetal_Muscle_Trunk.ChromatinAccessibility.H-24409.DS20242.twopass.merge150.wgt10.zgt2.wig
brain	SK-N-MC	SK_N_MC-DS14408.twopass.merge150.wgt10.zgt2.wig	fetal_muscle_trunk	Fetal_Muscle_Trunk	UW.Fetal_Muscle_Trunk.ChromatinAccessibility.H-24507.DS20544.twopass.merge150.wgt10.zgt2.wig
brain	SK-N-SH	SKNSH-DS8482.twopass.merge150.wgt10.zgt2.wig	fetal_muscle	Fetal_Muscle_Upper_Limb_Skeletal	UW.Fetal_Muscle_Upper_Limb_Skeletal.ChromatinAccessibility.H-23887.DS17661.twopass.merge150.wgt10.zgt2.wig
eye	WERI-Rb-1	WERI_Rb1-DS13681.twopass.merge150.wgt10.zgt2.wig	fetal_muscle_upper_trunk	Fetal_Muscle_Upper_Trunk	UW.Fetal_Muscle_Upper_Trunk.ChromatinAccessibility.H-23887.DS17664.twopass.merge150.wgt10.zgt2.wig
liver	8988T	wgEncodeOpenChromDnase8988tAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_placenta	Fetal_Placenta	UW.Fetal_Placenta.ChromatinAccessibility.H-23887.DS17639.twopass.merge150.wgt10.zgt2.wig
blood_vessel	AoSMC	wgEncodeOpenChromDnaseAosmcSerumfreeAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_placenta	Fetal_Placenta	UW.Fetal_Placenta.ChromatinAccessibility.H-23914.DS17744.twopass.merge150.wgt10.zgt2.wig
fetal_membrane	Chorion	wgEncodeOpenChromDnaseChorionAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_placenta	Fetal_Placenta	UW.Fetal_Placenta.ChromatinAccessibility.H-24272.DS19391.twopass.merge150.wgt10.zgt2.wig
blood	CLL	wgEncodeOpenChromDnaseCllAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_placenta	Fetal_Placenta	UW.Fetal_Placenta.ChromatinAccessibility.H-24409.DS20346.twopass.merge150.wgt10.zgt2.wig
skin	Fibrobl	wgEncodeOpenChromDnaseFibroblAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-23769.DS17307.twopass.merge150.wgt10.zgt2.wig
skin	FibroP	wgEncodeOpenChromDnaseFibropAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-23790.DS17387.twopass.merge150.wgt10.zgt2.wig
brain	Gliobla	wgEncodeOpenChromDnaseGlioblaAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-23833.DS17455.twopass.merge150.wgt10.zgt2.wig
blood	GM12891	wgEncodeOpenChromDnaseGm12891AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-23914.DS17756.twopass.merge150.wgt10.zgt2.wig
blood	GM12892	wgEncodeOpenChromDnaseGm12892AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-23941.DS17804.twopass.merge150.wgt10.zgt2.wig
blood	GM18507	wgEncodeOpenChromDnaseGm18507AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-24078.DS18428.twopass.merge150.wgt10.zgt2.wig
blood	GM19238	wgEncodeOpenChromDnaseGm19238AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-24259.DS19388.twopass.merge150.wgt10.zgt2.wig
blood	GM19239	wgEncodeOpenChromDnaseGm19239AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex	UW.Fetal_Renal_Cortex.ChromatinAccessibility.H-24493.DS20568.twopass.merge150.wgt10.zgt2.wig
blood	GM19240	wgEncodeOpenChromDnaseGm19240AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Left	UW.Fetal_Renal_Cortex_Left.ChromatinAccessibility.H-23871.DS17550.twopass.merge150.wgt10.zgt2.wig
es_cell	H9ES	wgEncodeOpenChromDnaseH9esAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Left	UW.Fetal_Renal_Cortex_Left.ChromatinAccessibility.H-24111.DS18542.twopass.merge150.wgt10.zgt2.wig
cervix	HeLa-S3	wgEncodeOpenChromDnaseHelas3fhs4hAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Left	UW.Fetal_Renal_Cortex_Left.ChromatinAccessibility.H-24169.DS18931.twopass.merge150.wgt10.zgt2.wig
liver	Hepatocytes	wgEncodeOpenChromDnaseHepatocytesAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Left	UW.Fetal_Renal_Cortex_Left.ChromatinAccessibility.H-24232.DS19257.twopass.merge150.wgt10.zgt2.wig
pancreatic_duct	HPDE6-E6E7	wgEncodeOpenChromDnaseHpde6e6e7AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Right	UW.Fetal_Renal_Cortex_Right.ChromatinAccessibility.H-23871.DS17545.twopass.merge150.wgt10.zgt2.wig
blastula	HTR8svn	wgEncodeOpenChromDnaseHtr8AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig	fetal_renal_cortex	Fetal_Renal_Cortex_Right	UW.Fetal_Renal_Cortex_Right.ChromatinAccessibility.H-24169.DS18928.twopass.merge150.wgt10.zgt2.wig

Tissue	Celltype	Annotation
liver	Huh-7.5	wgEncodeOpenChromDnaseHuh75AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
liver	Huh-7	wgEncodeOpenChromDnaseHuh7AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
ips_cell	iPS	wgEncodeOpenChromDnaseIpsAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
prostate	LNCaP	wgEncodeOpenChromDnaseLncapAndroAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
breast	MCF-7	wgEncodeOpenChromDnaseMcf7HypoxlacAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
brain	Medullo	wgEncodeOpenChromDnaseMedulloAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
skin	Melano	wgEncodeOpenChromDnaseMelanoAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
myometrium	Myometr	wgEncodeOpenChromDnaseMyometrAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
bone	Osteobl	wgEncodeOpenChromDnaseOsteoblAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
pancreas	PanIsletD	wgEncodeOpenChromDnasePanisdAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
pancreas	PanIslets	wgEncodeOpenChromDnasePanisletsAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
epithelium	pHTE	wgEncodeOpenChromDnasePhteAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
skin	ProgFib	wgEncodeOpenChromDnaseProgfibAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
liver	Stellate	wgEncodeOpenChromDnaseStellateAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
breast	T-47D	wgEncodeOpenChromDnaseT47dAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
urothelium	Urothelia	wgEncodeOpenChromDnaseUrotsaAlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
urothelium	Urothelia	wgEncodeOpenChromDnaseUrotsaUt189AlnAllReps.30000000.twopass.merge150.wgt10.zgt2.wig
embryonic_lung	WI-38	WI_38-DS14315.twopass.merge150.wgt10.zgt2.wig
embryonic_lung	WI-38	WI_38_TAM-DS14323.twopass.merge150.wgt10.zgt2.wig
breast	Breast_vHMEC	UW.Breast_vHMEC.ChromatinAccessibility.RM035.DS18406.twopass.merge150.wgt10.zgt2.wig
breast	Breast_vHMEC	UW.Breast_vHMEC.ChromatinAccessibility.RM035.DS18438.twopass.merge150.wgt10.zgt2.wig
blood	CD14_Primary_Cells	UW.CD14_Primary_Cells.ChromatinAccessibility.RO_01689.DS17391.twopass.merge150.wgt10.zgt2.wig
blood	CD14_Primary_Cells	UW.CD14_Primary_Cells.ChromatinAccessibility.RO_01701.DS17889.twopass.merge150.wgt10.zgt2.wig
blood	CD19_Primary_Cells	UW.CD19_Primary_Cells.ChromatinAccessibility.RO_01679.DS17186.twopass.merge150.wgt10.zgt2.wig
blood	CD19_Primary_Cells	UW.CD19_Primary_Cells.ChromatinAccessibility.RO_01689.DS17281.twopass.merge150.wgt10.zgt2.wig
blood	CD19_Primary_Cells	UW.CD19_Primary_Cells.ChromatinAccessibility.RO_01701.DS17440.twopass.merge150.wgt10.zgt2.wig
blood	CD20_Primary_Cells	UW.CD20_Primary_Cells.ChromatinAccessibility.RO_01778.DS17371.twopass.merge150.wgt10.zgt2.wig
blood	CD3_Primary_Cells	UW.CD3_Primary_Cells.ChromatinAccessibility.CB1_1-3-2011.DS17702.twopass.merge150.wgt10.zgt2.wig
blood	CD3_Primary_Cells	UW.CD3_Primary_Cells.ChromatinAccessibility.CB6_1-4-2011.DS17706.twopass.merge150.wgt10.zgt2.wig
blood	CD3_Primary_Cells	UW.CD3_Primary_Cells.ChromatinAccessibility.RO_01679.DS17198.twopass.merge150.wgt10.zgt2.wig
blood	CD3_Primary_Cells	UW.CD3_Primary_Cells.ChromatinAccessibility.RO_01701.DS17534.twopass.merge150.wgt10.zgt2.wig
blood	CD4_Primary_Cells	UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01549.DS15947.twopass.merge150.wgt10.zgt2.wig
blood	CD4_Primary_Cells	UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01679.DS17212.twopass.merge150.wgt10.zgt2.wig
blood	CD4_Primary_Cells	UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01689.DS17329.twopass.merge150.wgt10.zgt2.wig
blood	CD4_Primary_Cells	UW.CD4_Primary_Cells.ChromatinAccessibility.RO_01701.DS17881.twopass.merge150.wgt10.zgt2.wig
blood	CD56_Primary_Cells	UW.CD56_Primary_Cells.ChromatinAccessibility.RO_01679.DS17189.twopass.merge150.wgt10.zgt2.wig
blood	CD56_Primary_Cells	UW.CD56_Primary_Cells.ChromatinAccessibility.RO_01701.DS17443.twopass.merge150.wgt10.zgt2.wig
blood	CD8_Primary_Cells	UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01549.DS16012.twopass.merge150.wgt10.zgt2.wig
blood	CD8_Primary_Cells	UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01679.DS17203.twopass.merge150.wgt10.zgt2.wig
blood	CD8_Primary_Cells	UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01689.DS17332.twopass.merge150.wgt10.zgt2.wig
blood	CD8_Primary_Cells	UW.CD8_Primary_Cells.ChromatinAccessibility.RO_01701.DS17885.twopass.merge150.wgt10.zgt2.wig
fetal_adrenal_gland	Fetal_Adrenal_Gland	UW.Fetal_Adrenal_Gland.ChromatinAccessibility.H-22662.DS12528.twopass.merge150.wgt10.zgt2.wig
fetal_adrenal_gland	Fetal_Adrenal_Gland	UW.Fetal_Adrenal_Gland.ChromatinAccessibility.H-23769.DS17319.twopass.merge150.wgt10.zgt2.wig
fetal_adrenal_gland	Fetal_Adrenal_Gland	UW.Fetal_Adrenal_Gland.ChromatinAccessibility.H-24409.DS20343.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-22510.DS11872.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-22510.DS11877.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-22911.DS14464.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23266.DS14717.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23266.DS14718.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23284.DS14803.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23284.DS14815.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23399.DS15453.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-23548.DS16302.twopass.merge150.wgt10.zgt2.wig

Tissue	Celltype	Annotation
fetal_renal_cortex	Fetal_Renal_Cortex_Right	UW.Fetal_Renal_Cortex_Right.ChromatinAccessibility.H-24232.DS19254.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-23790.DS17381.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-23833.DS17451.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-23914.DS17760.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-24005.DS18088.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-24078.DS18431.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-24259.DS19386.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis	UW.Fetal_Renal_Pelvis.ChromatinAccessibility.H-24477.DS20448.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Left	UW.Fetal_Renal_Pelvis_Left.ChromatinAccessibility.H-23871.DS17553.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Left	UW.Fetal_Renal_Pelvis_Left.ChromatinAccessibility.H-24111.DS18666.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Left	UW.Fetal_Renal_Pelvis_Left.ChromatinAccessibility.H-24169.DS18964.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Left	UW.Fetal_Renal_Pelvis_Left.ChromatinAccessibility.H-24232.DS19238.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Right	UW.Fetal_Renal_Pelvis_Right.ChromatinAccessibility.H-23871.DS17548.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Right	UW.Fetal_Renal_Pelvis_Right.ChromatinAccessibility.H-24111.DS18663.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Right	UW.Fetal_Renal_Pelvis_Right.ChromatinAccessibility.H-24169.DS18961.twopass.merge150.wgt10.zgt2.wig
fetal_renal_pelvis	Fetal_Renal_Pelvis_Right	UW.Fetal_Renal_Pelvis_Right.ChromatinAccessibility.H-24232.DS19235.twopass.merge150.wgt10.zgt2.wig
fetal_skin	Fetal_Skin	UW.Fetal_Skin.ChromatinAccessibility.H-22337.DS10987.twopass.merge150.wgt10.zgt2.wig
fetal_spinal_cord	Fetal_Spinal_Cord	UW.Fetal_Spinal_Cord.ChromatinAccessibility.H-24111.DS18501.twopass.merge150.wgt10.zgt2.wig
fetal_spinal_cord	Fetal_Spinal_Cord	UW.Fetal_Spinal_Cord.ChromatinAccessibility.H-24409.DS20351.twopass.merge150.wgt10.zgt2.wig
fetal_spinal_cord	Fetal_Spinal_Cord	UW.Fetal_Spinal_Cord.ChromatinAccessibility.H-24493.DS20530.twopass.merge150.wgt10.zgt2.wig
fetal_spleen	Fetal_Spleen	UW.Fetal_Spleen.ChromatinAccessibility.H-23399.DS17448.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23589.DS16530.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23758.DS17172.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23769.DS17325.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23887.DS17659.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23914.DS17750.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-23964.DS17878.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-24005.DS17963.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-24078.DS18389.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-24125.DS18821.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-24401.DS20349.twopass.merge150.wgt10.zgt2.wig
fetal_stomach	Fetal_Stomach	UW.Fetal_Stomach.ChromatinAccessibility.H-24507.DS20546.twopass.merge150.wgt10.zgt2.wig
fetal_testes	Fetal_Testes	UW.Fetal_Testes.ChromatinAccessibility.H24198_H23640_H24042.DS18942.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23589.DS16490.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23663.DS16841.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23769.DS17323.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23833.DS17474.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23964.DS17875.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-23964.DS17876.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-24078.DS18382.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-24259.DS19287.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-24401.DS20335.twopass.merge150.wgt10.zgt2.wig
fetal_thymus	Fetal_Thymus	UW.Fetal_Thymus.ChromatinAccessibility.H-24409.DS20341.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Abdomen	UW.Fibroblasts_Fetal_Skin_Abdomen.ChromatinAccessibility.H-24259.DS19558.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Abdomen	UW.Fibroblasts_Fetal_Skin_Abdomen.ChromatinAccessibility.H-24259.DS19561.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Back	UW.Fibroblasts_Fetal_Skin_Back.ChromatinAccessibility.H-24218.DS19233.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Biceps_Left	UW.Fibroblasts_Fetal_Skin_Biceps_Left.ChromatinAccessibility.H-24259.DS19857.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Biceps_Left	UW.Fibroblasts_Fetal_Skin_Biceps_Left.ChromatinAccessibility.H-24259.DS19867.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Biceps_Right	UW.Fibroblasts_Fetal_Skin_Biceps_Right.ChromatinAccessibility.H-24259.DS19745.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Biceps_Right	UW.Fibroblasts_Fetal_Skin_Biceps_Right.ChromatinAccessibility.H-24259.DS19761.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Quadriceps_Left	UW.Fibroblasts_Fetal_Skin_Quadriceps_Left.ChromatinAccessibility.H-24259.DS20046.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Quadriceps_Left	UW.Fibroblasts_Fetal_Skin_Quadriceps_Left.ChromatinAccessibility.H-24259.DS20056.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Quadriceps_Right	UW.Fibroblasts_Fetal_Skin_Quadriceps_Right.ChromatinAccessibility.H-24259.DS19943.twopass.merge150.wgt10.zgt2.wig

Tissue	Celltype	Annotation
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-24279.DS20221.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-24297.DS20226.twopass.merge150.wgt10.zgt2.wig
fetal_brain	Fetal_Brain	UW.Fetal_Brain.ChromatinAccessibility.H-24381.DS20231.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-22662.DS12531.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-22727.DS12810.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23468.DS15839.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23500.DS16018.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23524.DS16146.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23589.DS16500.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23604.DS16582.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23617.DS16621.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23663.DS16819.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-23744.DS19431.twopass.merge150.wgt10.zgt2.wig
fetal_heart	Fetal_Heart	UW.Fetal_Heart.ChromatinAccessibility.H-24042.DS19427.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23500.DS16027.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23524.DS16164.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23604.DS16563.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23744.DS17094.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23758.DS17157.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23769.DS17313.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23808.DS17422.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23833.DS17462.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23855.DS17502.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23887.DS17647.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23914.DS17748.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23941.DS17785.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-23964.DS17841.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-24005.DS17990.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_large	Fetal_Intestine_Large	UW.Fetal_Intestine_Large.ChromatinAccessibility.H-24111.DS18499.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23604.DS16559.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23640.DS16712.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23663.DS16822.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23724.DS16975.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23744.DS17092.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23758.DS17150.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23769.DS17317.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23808.DS17425.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23864.DS17844.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23887.DS17643.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23914.DS17763.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-23941.DS17808.twopass.merge150.wgt10.zgt2.wig
fetal_intestine,_small	Fetal_Intestine_Small	UW.Fetal_Intestine_Small.ChromatinAccessibility.H-24111.DS18495.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-22337.DS10986.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-22676.DS12635.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-23524.DS16139.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-23663.DS16837.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-23855.DS17522.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-23914.DS17753.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney	UW.Fetal_Kidney.ChromatinAccessibility.H-24507.DS20564.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney_Left	UW.Fetal_Kidney_Left.ChromatinAccessibility.H-23589.DS16446.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney_Left	UW.Fetal_Kidney_Left.ChromatinAccessibility.H-23604.DS16579.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney_Left	UW.Fetal_Kidney_Left.ChromatinAccessibility.H-23640.DS16805.twopass.merge150.wgt10.zgt2.wig
fetal_kidney	Fetal_Kidney_Left	UW.Fetal_Kidney_Left.ChromatinAccessibility.H-23758.DS17140.twopass.merge150.wgt10.zgt2.wig

Tissue	Celltype	Annotation
fibroblast	Fibroblasts_Fetal_Skin_Quadriceps_Right	UW.Fibroblasts_Fetal_Skin_Quadriceps_Right.ChromatinAccessibility.H-24259.DS19948.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Scalp	UW.Fibroblasts_Fetal_Skin_Scalp.ChromatinAccessibility.H-24259.DS19444.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Scalp	UW.Fibroblasts_Fetal_Skin_Scalp.ChromatinAccessibility.H-24259.DS19449.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Upper_Back	UW.Fibroblasts_Fetal_Skin_Upper_Back.ChromatinAccessibility.H-24259.DS19696.twopass.merge150.wgt10.zgt2.wig
fibroblast	Fibroblasts_Fetal_Skin_Upper_Back	UW.Fibroblasts_Fetal_Skin_Upper_Back.ChromatinAccessibility.H-24259.DS19706.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_BMP4_Derived_Mesendoderm_Cultured_Cells	UW.H1_BMP4_Derived_Mesendoderm_Cultured_Cells.ChromatinAccessibility.DS18732.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_BMP4_Derived_Mesendoderm_Cultured_Cells	UW.H1_BMP4_Derived_Mesendoderm_Cultured_Cells.ChromatinAccessibility.DS19310.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_BMP4_Derived_Trophoblast_Cultured_Cells	UW.H1_BMP4_Derived_Trophoblast_Cultured_Cells.ChromatinAccessibility.DS18736.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_BMP4_Derived_Trophoblast_Cultured_Cells	UW.H1_BMP4_Derived_Trophoblast_Cultured_Cells.ChromatinAccessibility.DS19317.twopass.merge150.wgt10.zgt2.wig
es_cell	H1	UW.H1.ChromatinAccessibility.DS18873.twopass.merge150.wgt10.zgt2.wig
es_cell	H1	UW.H1.ChromatinAccessibility.DS19100.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_Derived_Mesenchymal_Stem_Cells	UW.H1_Derived_Mesenchymal_Stem_Cells.ChromatinAccessibility.DS20671.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_Derived_Mesenchymal_Stem_Cells	UW.H1_Derived_Mesenchymal_Stem_Cells.ChromatinAccessibility.DS21042.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_Derived_Neuronal_Progenitor_Cultured_Cells	UW.H1_Derived_Neuronal_Progenitor_Cultured_Cells.ChromatinAccessibility.DS18739.twopass.merge150.wgt10.zgt2.wig
es_cell	H1_Derived_Neuronal_Progenitor_Cultured_Cells	UW.H1_Derived_Neuronal_Progenitor_Cultured_Cells.ChromatinAccessibility.DS20153.twopass.merge150.wgt10.zgt2.wig
es_cell	H9	UW.H9.ChromatinAccessibility.DS18517.twopass.merge150.wgt10.zgt2.wig
es_cell	H9	UW.H9.ChromatinAccessibility.DS18522.twopass.merge150.wgt10.zgt2.wig
lung	IMR90	UW.IMR90.ChromatinAccessibility.DS11759.twopass.merge150.wgt10.zgt2.wig
lung	IMR90	UW.IMR90.ChromatinAccessibility.DS11764.twopass.merge150.wgt10.zgt2.wig
lung	IMR90	UW.IMR90.ChromatinAccessibility.DS13219.twopass.merge150.wgt10.zgt2.wig
lung	IMR90	UW.IMR90.ChromatinAccessibility.DS13229.twopass.merge150.wgt10.zgt2.wig
ips_cell	iPS_DF_19.11	UW.iPS_DF_19.11.ChromatinAccessibility.DS15153.twopass.merge150.wgt10.zgt2.wig
ips_cell	iPS_DF_19.7	UW.iPS_DF_19.7.ChromatinAccessibility.DS15148.twopass.merge150.wgt10.zgt2.wig
ips_cell	iPS_DF_4.7	UW.iPS_DF_4.7.ChromatinAccessibility.DS15169.twopass.merge150.wgt10.zgt2.wig
ips_cell	iPS_DF_6.9	UW.iPS_DF_6.9.ChromatinAccessibility.DS15164.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_00738.DS13199.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01480.DS12771.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01492.DS12774.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01492.DS14206.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01508.DS11202.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01517.DS14129.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01520.DS12785.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01527.DS11666.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01535.DS12274.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01535.DS14197.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01536.DS12339.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01549.DS12734.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01562.DS13196.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD34_Primary_Cells	UW.Mobilized_CD34_Primary_Cells.ChromatinAccessibility.RO_01701.DS17112.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD4_Primary_Cells	UW.Mobilized_CD4_Primary_Cells.ChromatinAccessibility.RO_01679.DS16955.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD4_Primary_Cells	UW.Mobilized_CD4_Primary_Cells.ChromatinAccessibility.RO_01701.DS17175.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD56_Primary_Cells	UW.Mobilized_CD56_Primary_Cells.ChromatinAccessibility.RO_01689.DS16376.twopass.merge150.wgt10.zgt2.wig
blood	Mobilized_CD8_Primary_Cells	UW.Mobilized_CD8_Primary_Cells.ChromatinAccessibility.RO_01679.DS16962.twopass.merge150.wgt10.zgt2.wig
fibroblast	Penis_Foreskin_Fibroblast_Primary_Cells	UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin01.DS18224.twopass.merge150.wgt10.zgt2.wig
fibroblast	Penis_Foreskin_Fibroblast_Primary_Cells	UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin01.DS18229.twopass.merge150.wgt10.zgt2.wig
fibroblast	Penis_Foreskin_Fibroblast_Primary_Cells	UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin02.DS18252.twopass.merge150.wgt10.zgt2.wig
fibroblast	Penis_Foreskin_Fibroblast_Primary_Cells	UW.Penis_Foreskin_Fibroblast_Primary_Cells.ChromatinAccessibility.skin02.DS18256.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Keratinocyte_Primary_Cells	UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18692.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Keratinocyte_Primary_Cells	UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18695.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Keratinocyte_Primary_Cells	UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin02.DS18714.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Keratinocyte_Primary_Cells	UW.Penis_Foreskin_Keratinocyte_Primary_Cells.ChromatinAccessibility.skin02.DS18718.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Melanocyte_Primary_Cells	UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18590.twopass.merge150.wgt10.zgt2.wig
skin	Penis_Foreskin_Melanocyte_Primary_Cells	UW.Penis_Foreskin_Melanocyte_Primary_Cells.ChromatinAccessibility.skin01.DS18601.twopass.merge150.wgt10.zgt2.wig

Supplementary Table 2-2. 95 SLE risk SNPs for eQTL analyses and for GRS calculation.

SNP	chr	pos	Reference_gene	Risk.allele	Alt.allele	OR
rs4649203	1	24519920	*IFNLR1_LOC284632*	G	A	1.16
rs2476601	1	114377568	*PTPN22*	A	G	1.35
rs1801274	1	161479745	*FCGR2A*	G	A	1.16
rs2205960	1	173191475	*TNFSF4_LOC100506023*	T	G	1.46
rs1418190	1	173361979	*LOC100506023*	T	C	1.23
rs17849502	1	183532580	*NCF2*	T	G	2.57
rs10911628	1	184649503	*C1orf21_EDEM3*	A	C	1.95
rs34889541	1	198594769	*ATP6V1G3_PTPRC*	G	A	1.23
rs2297550	1	206643772	*IKBKE*	G	C	1.16
rs3024505	1	206939904	*MAPKAPK2_IL10*	A	G	1.19
rs9782955	1	236039877	*LYST*	C	T	1.16
rs7579944	2	30445026	*SNORA10B_LBH*	C	T	1.13
rs17321999	2	30479857	*LBH*	C	A	1.20
rs13385731	2	33701890	*RASGRP3*	T	C	1.43
rs6740462	2	65667272	*SPRED2_MIR4778*	A	C	1.10
rs6705628	2	74208362	*DGUOK-AS1*	C	T	1.33
rs2111485	2	163110536	*FAP_IFIH1*	G	A	1.15
rs3821236	2	191902758	*STAT4*	A	G	1.49
rs11889341	2	191943742	*STAT4*	T	C	1.73
rs3768792	2	213871709	*IKZF2*	G	A	1.24
rs6445972	3	58321707	*PXK*	T	C	1.23
rs1132200	3	119150836	*TMEM39A*	C	T	1.39
rs2222631	3	119272391	*CD80*	A	G	1.09
rs564799	3	159728987	*IL12A-AS1*	C	T	1.14
rs10936599	3	169492101	*MYNN*	C	T	1.14
rs6762714	3	188470238	*LPP*	T	C	1.16
rs4690229	4	970724	*DGKQ_SLC26A1*	T	A	1.13
rs340630	4	87958395	*AFF1*	A	G	1.21
rs10028805	4	102737250	*BANK1*	G	A	1.20
rs907715	4	123535053	*IL21*	C	T	1.16
rs7726159	5	1282319	*TERT*	A	C	1.21
rs7726414	5	133431834	*VDAC1_TCF7*	T	C	1.45
rs7708392	5	150457485	*TNIP1*	C	G	1.27
rs2421184	5	158886939	*LINC01845*	A	G	1.19
rs2431697	5	159879978	*PTTG1_MIR3142HG*	T	C	1.45
rs17603856	6	16630898	*ATXN1*	T	G	1.14
rs36014129	6	25884519	*SLC17A3_SLC17A2*	A	G	1.50
rs597325	6	91002494	*BACH2*	G	A	1.12
rs6568431	6	106588806	*PRDM1_ATG5*	A	C	1.20
rs6932056	6	138242437	*TNFAIP3_PERP*	C	T	1.83
rs2327832	6	137973068	*OLIG3_LOC102723649*	G	A	1.22

SNP	chr	pos	Reference_gene	Risk.allele	Alt.allele	OR
rs849142	7	28185891	JAZF1	T	C	1.19
rs4917014	7	50305863	C7orf72_IKZF1	T	G	1.39
rs73135369	7	73940978	GTF2IRD1	C	T	1.32
rs1167796	7	75173180	HIP1	G	A	1.20
rs4728142	7	128573967	KCP_IRF5	A	G	1.43
rs2070197	7	128589000	IRF5	C	T	1.88
rs2980512	8	8140901	FAM86B3P_PRAG1	C	T	1.15
rs2736340	8	11343973	FAM167A_BLK	T	C	1.35
rs7829816	8	56849386	LYN	A	G	1.30
rs1966115	8	79556891	PKIA_ZC2HC1A	A	G	1.14
rs1887428	9	4984530	JAK2	G	C	1.16
rs7097397	10	50025396	WDFY4	G	A	1.30
rs877819	10	50042951	WDFY4	A	G	1.46
rs4948496	10	63805617	ARID5B	C	T	1.18
rs4917385	10	105003721	NT5C2_RPEL1	G	T	1.39
rs12802200	11	566936	MIR210HG	C	A	1.23
rs2732552	11	35084592	PDHX_LOC100507144	C	T	1.22
rs494003	11	65542298	AP5B1	A	G	1.14
rs3794060	11	71187679	NADSYN1	C	T	1.23
rs4639966	11	118573519	TREH_DDX6	C	T	1.29
rs6590330	11	128311059	LINC02098_ETS1	A	G	1.37
rs7941765	11	128499000	ETS1_LOC101929538	C	T	1.14
rs12822507	12	12773521	CREBL2	A	G	1.16
rs34330	12	12870695	CDKN1B	C	T	1.19
rs10506216	12	43130885	LINC02450_LINC02461	A	G	1.67
rs4622329	12	102321935	DRAM1_WASHC3	A	G	1.19
rs10774625	12	111910219	ATXN2	A	G	1.13
rs1059312	12	129278864	SLC15A4	G	A	1.17
rs8016947	14	35832666	PSMA6_NFKBIA	G	T	1.12
rs4902562	14	68731458	RAD51B	A	G	1.14
rs11073328	15	38764843	FAM98B	T	C	1.94
rs34933034	15	75079474	CSK	A	G	1.32
rs2289583	15	75311036	SCAMP5	A	C	1.19
rs8023715	15	97607681	SPATA8_LINC02254	A	C	1.81
rs9652601	16	11174365	CLEC16A	G	A	1.21
rs16972959	16	23901376	PRKCB	A	G	1.23
rs7197475	16	30642867	ZNF689_PRR14	T	C	1.31
rs34572943	16	31272353	ITGAM	A	G	1.71
rs223881	16	57386566	PLLP_CCL22	T	C	1.15
rs1170426	16	68603798	ZFP90	C	T	1.12
rs2934498	16	85968282	IRF8_LINC01082	G	A	1.25
rs2280381	16	86018633	IRF8_LINC01082	T	C	1.16

SNP	chr	pos	Reference_gene	Risk.allele	Alt.allele	OR
rs2286672	17	4712617	PLD2	T	C	1.25
rs2941509	17	37921194	IKZF3	T	C	1.35
rs930297	17	73404537	MIR3678_TMEM94	T	C	1.21
rs1610555	18	67543147	CD226	T	G	1.19
rs3093030	19	10397403	ICAM4_ICAM1	T	C	1.16
rs2304256	19	10475652	TYK2	C	A	1.24
rs2305772	19	52033742	SIGLEC6	G	A	1.16
rs4810485	20	44747947	CD40	T	G	1.59
rs11697848	20	48575315	RNF114_SNAI1	T	C	2.12
rs463426	22	21809185	HIC2_TMEM191C	T	C	1.28
rs7444	22	21976934	UBE2L3	C	T	1.27
rs61616683	22	39755773	SYNGR1	T	C	1.27

Notes: Alt.allele = alternative allele. OR = Odds Ratio.

Appendix B. Chapter 5 Supplementary Tables

Supplementary Table 5-1: SLE Renal GWAS in SLE main Cohort (35 Suggestive regions, P ≤ 1E-05)

			Alleles		*P* value			Annotation	
Suggetive Regions	**tagSNP***	**Position**	**Reference**	**Alternative**	**SLE main cohort**	**SLEGEN cohort**	**Meta-analysis of two cohorts**	**Function**	**Reference Gene**
1p36.12	rs10607584	21678940	GGA	G	6.17742E-06	0.791527	1.29379E-05	intergenic	*ECE1;NBPF3*
1q31.3	rs9427906	195926095	T	C	4.89737E-06	0.558663	6.32048E-05	intergenic	*NONE;LINC01724*
1q41	rs11572579	217011607	C	T	4.39503E-06	0.0985657	0.011945786	intronic	*ESRRG*
1q42.2	rs57268081	230866723	A	C	3.12406E-06	0.3431	0.000215125	intergenic	*AGT;CAPN9*
2q24.3	rs79968435	168051427	T	C	7.10432E-06	0.856816	3.92125E-05	intronic	*XIRP2*
3p14.1	rs34242854	68061272	C	G	6.08601E-06	0.672348	1.1343E-05	intronic	*FAM19A1*
3q12.3	rs141037434	101966528	A	G	4.80943E-06	0.340595	0.827584767	intergenic	*LOC101929411;ZPLD1*
3q25.2	rs35259387	152805860	G	T	6.67233E-06	0.0230238	0.000479802	intergenic	*P2RY1;RAP2B*
3q26.2	rs2421072	168667440	C	T	0.000006047	0.978959	2.71723E-05	intergenic	*LINC02082;MECOM*
4p14	rs36185839	37312869	C	T	6.76737E-06	0.386835	7.25473E-06	intronic	*NWD2*
4q27	rs62324163	123501681	T	G	3.68549E-06	0.270292	2.52989E-06	intergenic	*IL2;IL21*
4q34.1	rs72698624	176025380	T	C	9.08584E-06	0.807888	5.59577E-05	intergenic	*ADAM29;GPM6A*
5p12	rs148454819	43585130	G	A	7.39295E-06	0.093852	4.70122E-05	ncRNA_intronic	*NNT-AS1*
5q13.3	rs77835099	73573014	C	T	3.59984E-06	0.406199	0.198502641	intergenic	*ARHGEF28;LINC01335*
6q14.1	rs117308968	83686996	T	A	1.60468E-06	0.255017	3.21595E-05	intronic	*UBE3D*
7q21.3	rs45499693	93067597	A	C	8.0478E-06	0.433873	0.001278023	intronic	*CALCR*
8p12	rs183396989	29161973	G	A	4.4386E-06	0.777606	0.856533823	intergenic	*KIF13B;DUSP4*
8q24.3	rs55883233	141463613	C	CCATTCTGCATATG CAGAGCCCACGA	3.64049E-06	0.37963	3.54305E-06	intronic	*TRAPPC9*
9q33.3	rs576074053	130262732	C	T	8.4443E-07	0.756719	0.752306379	intronic	*LRSAM1*
11q21	rs35054871	92823118	CT	C	2.2403E-06	0.7869	6.6325E-06	intergenic	*MTNR1B;SLC36A4*
12q14.1	rs61941637	61818040	A	T	7.67775E-06	0.933739	3.20274E-05	intergenic	*SLC16A7;FAM19A2*
12q23.3	rs7957648	104954189	T	C	8.17982E-06	0.196864	0.000147574	intronic	*CHST11*
13q12.12	rs59543007	23710017	G	GA	6.15852E-06	0.799654	4.57152E-05	intergenic	*LINC00621;SGCG*
13q14.11	rs11372973	41612202	T	TA	9.00371E-06	0.399211	0.000147939	intergenic	*ELF1;WBP4*
14q23.2	rs1271565	64044097	T	C	4.57017E-06	0.886522	1.80464E-05	intergenic	*PPP2R5E;WDR89*
14q24.1	rs4899257	69253984	G	T	4.61264E-06	0.281105	3.14879E-06	downstream	*ZFP36L1*
14q31.1	rs4904052	82755793	A	G	7.54394E-06	0.0355177	0.013204381	intergenic	*LINC02311;LINC02301*
15q21.1	rs112366266	48800870	G	A	7.73387E-06	0.439979	0.475934483	exonic	*FBN1*
15q26.1	rs8035491	90342353	G	A	5.9188E-06	0.545431	5.92267E-05	intronic	*ANPEP*
17q21.2	rs191116624	39042992	T	C	3.3014E-06	0.283989	0.004173173	intergenic	*KRT20;KRT23*
18q11.2	rs2226706	21650964	T	C	9.50537E-06	0.488401	1.4465E-05	intronic	*TTC39C*
18q22.2	rs561244604	67151529	A	C	1.48608E-06	0.349488	0.902974114	intronic	*DOK6*
20p12.3	rs16991615	5948227	G	A	7.2717E-06	0.520827	1.73527E-05	exonic	*MCM8*
22q13.2	rs191312901	43757297	G	A	4.54433E-06	0.703811	0.002464557	intergenic	*SCUBE1;LINC01639*
Xp21.1	rs2180203	32407001	A	T	4.02123E-06	0.141587	1.75129E-06	intronic	*DMD*

Supplementary Table 5-2. SLE Renal GWAS in SLEGEN Cohort (23 Suggestive regions, P ≤ 1E-05)

			Alleles		*P* value			Annotation	
Suggetive Regions	**tagSNP***	**Position**	**Reference**	**Alternative**	**SLE main cohort**	**SLEGEN cohort**	**Meta-analysis of two cohorts**	**Function**	**Reference Gene**
1p33	rs3034530	48275827	C	T	0.578785	8.20143E-06	0.030747717	intronic	TRABD2B
3q29	rs71921811	192781673	G	GT	0.0155719	8.06823E-06	0.000100578	intergenic	MB21D2;HRASLS
4q26	rs200625880	120021625	TAA	T	0.713795	9.55342E-07	0.045502639	intergenic	SYNPO2;MYOZ2
5q14.2	rs76400304	82004240	G	C	0.704575	9.27711E-06	0.698403444	intergenic	ATP6AP1L;MIR3977
5q35.1	rs111947887	172767169	G	GA	0.945642	2.62665E-06	0.124481492	intergenic	STC2;MIR8056
5q35.1	rs74411785	172773769	T	C	0.714371	6.12468E-06	0.203289887	upstream	MIR8056
5q35.1	rs75428302	172775585	CG	C	0.78758	7.86472E-06	0.199563186	intergenic	MIR8056;LOC285593
7q31.1	rs201305724	108287113	TTAATA	T	0.298926	8.70395E-06	0.310763925	intergenic	DNAJB9;C7orf66
8p22	rs151276961	18034539	A	C	0.00943098	3.83641E-06	0.148239853	intronic	NAT1
8q12.2	rs4289856	61813615	G	A	0.231764	9.21871E-06	0.861680027	intergenic	CHD7;LOC100130298
9p13.3	rs10814291	35888258	G	A	0.95064	5.46102E-06	0.094898991	intergenic	OR13J1;HRCT1
10p12.1	rs111595213	25432752	C	CA	0.414948	9.69865E-06	0.924353509	ncRNA_intronic	LINC01516
10q26.13	rs17612814	123255847	G	A	0.921274	6.94594E-06	0.142036482	intronic	FGFR2
10q26.13	rs78233824	123383396	G	A	0.790281	7.23021E-06	0.059657749	intergenic	FGFR2;ATE1
11q22.1	rs6589875	98928962	T	A	0.721736	2.13118E-06	0.219998653	intronic	CNTN5
14q22.3	rs375958485	56514369	T	TTTATATATA	0.552054	8.71121E-06	0.631190004	intergenic	LINC00520;PELI2
14q31.1	rs34242066	80292343	A	AT	0.372434	5.11099E-06	0.024380592	intronic	NRXN3
16p13.3	rs1684584	4758465	C	A	0.502354	6.46877E-06	0.034542246	intronic	ANKS3
16p13.3	rs859315	4785807	G	A	0.84269	3.96597E-06	0.086980447	intronic	C16orf71
16p13.3	rs12162062	4800009	T	C	0.555661	7.64271E-06	0.037843641	UTR3	ZNF500
16q23.2	rs11150231	79956229	A	C	0.080014	8.40555E-06	0.001353785	intergenic	LINC01229;LOC102724084
18q23	rs113566940	76381494	T	C	0.897563	4.0874E-06	0.131387702	intergenic	LINC01029;SALL3
21q22.11	rs13047194	32835057	T	G	0.71495	7.24242E-06	0.207079984	intronic	TIAM1

Supplementary Table 5-3. Meta-analysis of SLE Renal GWAS of SLE main cohort and SLEGEN Cohort (21 Suggestive regions, P ≤ 1E-05)

			Alleles		*P* value				Annotation
Suggetive Regions	tagSNP	Position	Reference	Alternative	SLE main cohort	SLEGEN cohort	Meta-analysis of two cohorts	Function	Reference Gene
1p36.13	rs4920553	19163656	G	A	2.10452E-05	0.092638	4.24613E-06	intergenic	PAX7;TAS1R2
1q22	rs10157319	156157258	C	T	0.000022706	0.0519644	6.66881E-06	intergenic	SEMA4A;SLC25A44
1q22	rs2842879	156163695	G	T	0.000032893	0.0525782	9.45566E-06	upstream	SLC25A44
2p23.3	rs142186512	26400413	AT	A	0.000310971	0.00108808	6.41531E-06	intronic	GAREM2
2p23.3	rs116300430	26444690	A	G	0.00052792	0.000863644	9.48627E-06	intronic	HADHA
2p23.3	rs143543307	26595574	A	G	0.000171023	0.000457101	1.88256E-06	intronic	SELENOI
2q33.1	rs16831252	200170530	T	C	0.000171036	0.00278706	3.75403E-06	intronic	SATB2
3p14.1	rs1374837	68065201	C	G	5.27017E-06	0.660839	9.65206E-06	intronic	FAM19A1
4p14	rs36185839	37312869	C	T	6.76737E-06	0.386835	7.25473E-06	intronic	NWD2
4q27	rs62324163	123501681	T	G	3.68549E-06	0.270292	2.52989E-06	intergenic	IL2;IL21
4q34.1	rs1600330	176033852	T	C	1.42577E-06	0.959344	8.05694E-06	intergenic	ADAM29;GPM6A
5q22.2	rs72797840	111839996	A	G	4.32864E-05	0.0914914	8.81893E-06	intergenic	EPB41L4A-AS2;LINC02200
6q21	rs6935764	105903329	C	T	0.000175699	0.0070327	7.60639E-06	intergenic	PREP;PRDM1
8q24.3	rs55883233	141463613	C	CCATTCTGCATATGCA GAGCCCACGA	3.64049E-06	0.37963	3.54305E-06	intronic	TRAPPC9
11q21	rs35054871	92823118	CT	C	2.2403E-06	0.7869	6.6325E-06	intergenic	MTNR1B;SLC36A4
13q14.2	rs117874987	47454497	C	T	1.00164E-05	0.0462432	7.60529E-06	intronic	HTR2A
14q24.1	rs4899257	69253984	G	T	4.61264E-06	0.281105	3.14879E-06	downstream	ZFP36L1
18q11.2	rs72196202	21750898	CACTGT	C	3.09364E-06	0.728187	8.16927E-06	intronic	OSBPL1A
18q21.32	rs373465415	57128168	AAAAT	A	4.58707E-05	0.0129916	4.08426E-06	intronic	CCBE1
21q22.13	rs2835447	38090161	T	C	4.41334E-05	0.0224126	3.90073E-06	intronic	SIM2
Xp21.1	rs2144498	32406349	G	A	2.66455E-05	0.125521	9.62627E-06	intronic	DMD

Appendix C. Chapter 6 Supplementary Protocols

Date Written: 14/02/14 **Version:** 5.0
Amended: 09/02/2015_Lingyan Chen

PBMC Full Immunophenotype Staining for Flow Cytometry

Reagents Needed:
1x PBS
2% PBS/BSA [200ml PBS + 4g BSA (store in 4^{o}C)]
Histopaque [*Sigma*]
3x Stabilizing Fixative [*BD Biosciences*]
Permeabilization Buffer [PBS (49ml), 2% FBS (1ml), 0.1% Triton-X100 (50ul)]
Human TruStain FcX [*Biolegend*]
Live/Dead Blue Fixable Viability Dye [*Life Technologies*]

Antibodies Needed:
See document "Full Staining Panel.xlsx".

Notes

- All antibodies are kept at 4^{o}C, with the exception of the IKZF1 antibody and the Alexa Fluor 488 anti-goat IgG F(ab')2 fragment which are kept at -80 and -20 respectively. Thaw the IKZF1 antibody on ice prior to staining. The secondary antibody is diluted in glycerol, so will not need to be thawed first.
- Buffers containing FBS should be made fresh before each experiment.
- This protocol assumes a single blood sample is being stained. Adjust accordingly for additional samples.
- g(RCF- relative centrifugal force)

A. PBMC Isolation

1. Label flow cytometry tubes for each staining panel, crossed with each IKZF antibody, as well as for the IKZF Isotype control, viability dye single stained and an unstained sample. For a single blood sample, this will yield 18 tubes (4 ZF x 4 Panel, 1 US, 1 VDSS).

US	VD SS		
Panel_1 & IKZF1	Panel_1 & IKZF2	Panel_1 & IKZF3	Panel_1 & IKZF IC
Panel_2 & IKZF1	Panel_2 & IKZF2	Panel_2 & IKZF3	Panel_2 & IKZF IC
Panel_3 & IKZF1	Panel_3 & IKZF2	Panel_3 & IKZF3	Panel_3 & IKZF IC
Panel_4 & IKZF1	Panel_4 & IKZF2	Panel_4 & IKZF3	Panel_4 & IKZF IC

2. Add 15ml of Histopaque to the SepMate-50 tube, pipetting through the hole in the divider.
   (*The Histopaque should be added to the SepMate-50 tube 30min prior to the collecting of samples. Keep it at room temperature in the dark. Usually, each patient sample needs 4 tubes of Histopaque, and each control sample needs 2 tubes of Histopaque.)

3. Dilute the blood with an equal volume of PBS.
   (*Usually, each patient sample needs two 50ml tubes to dilute ~40ml whole blood, and each control sample needs one 50ml tubes to dilute ~20ml whole blood.)

4. Pipette the blood slowly into the SepMate tube, avoiding it passing through the hole in the divider.
   (*20ml diluted whole blood can be added into each 15ml Histopaque.)

5. Spin the blood sample for 10 minutes at room temperature at 2000xg, with brake.

6. Decant the plasma and PBMC portion into a new tube quickly and add equal volume of 2% PBS/BSA to each tube (up to 40ml for each tube), spin at 120 xg for 10 minutes at room temperature, with the **brake off.**

7. Aspirate the supernatant from the cells, and suspend the cells in 2ml 2%PBS/BSA for each tube and combine them into one tube. Add 2%PBS/BSA to 40ml.

8. Spin at room temperature for 8minutes at 300 xg.

9. Aspirate the supernatant from the cells, and resuspend in 10ml PBS. Put the cells on ice.

10. Count the cells, then aliquot 10^6 cells into each tube and place on ice. Add 1ml PBS to each tube, and then spin at 350xg for 5 minutes at 4^oC.

    (*count cells: 2ul cell solution + 18ul trypan blue. Number of cells/ml = $X/4*10^5$)

B. Surface Marker Staining and Fixation

11. While the samples are in the centrifuge, bring the viability dye to room temperature. If the dye has not yet been reconstituted, then add 40ul DMSO and mix well with the pipette.
    (* Turn upside-down to mix the solution)

12. Prepare a master mix for the viability dye. In a 50ml tube, mix 20ml PBS with 20ul of viability dye.

13. Decant the supernatant from the flow tubes by inversion and place them in a tube rack at room temperature.

14. Resuspend the cells in 1ml of the viability dye master mix (* Vortex for 1sec). Place them at room temperature in the dark for 30 minutes. Add PBS to the US sample.

15. Wash the cells in 1ml PBS and spin at 350xg for 5 minutes at 4°C.

16. Resuspend the cells in 50ul 2% PBS/BSA. **Add 5ul of TruStain FcX to each sample, including the US and VDSS samples, mix (*gently flick the bottom of the tubes to mix) and incubate in the dark for 10 minutes at room temperature.**

17. Add the following four antibodies to the appropriate tubes, with the following volumes of antibody per tube – 5ul CCR7 BV412 (Panel 1), 5ul CXCR3 BV605 (Panel 1), 2ul CCR6 PE/Cy7 (Panel 1), 1ul CCR4 APC (Panel 2).

18. Place all tubes in a 37°C incubator for 15 minutes.

19. During these 15 minutes, prepare the 4 surface stain master mixes using 2% PBS/BSA as shown in **Appendix 1**.

Panel_1 & IKZF1	Panel_1 & IKZF2	Panel_1 & IKZF3	Panel_1 & IKZF IC
Panel_2 & IKZF1	**Panel_2 & IKZF2**	**Panel_2 & IKZF3**	**Panel_2 & IKZF IC**
Panel_3 & IKZF1	Panel_3 & IKZF2	Panel_3 & IKZF3	Panel_3 & IKZF IC
Panel_4 & IKZF1	**Panel_4 & IKZF2**	**Panel_4 & IKZF3**	**Panel_4 & IKZF IC**

20. Remove tubes from the incubator. Add **50ul** of the appropriate master mix (*gently flick the bottom of the tubes to mix) to the tubes (Panels 1-4). Add 50ul 2% PBS/BSA to the **US and VD SS samples.**

21. **To the four ZF1 tubes, add 5ul of the PE isotype control. To ZF2, ZF3 and the ZF IC tubes, add 5ul of OX40L PE**. Incubate at room temperature for 20 minutes in the dark.

PE isotype	**OX40L PE**		
Panel_1 & IKZF1	**Panel_1 & IKZF2**	**Panel_1 & IKZF3**	**Panel_1 & IKZF IC**
Panel_2 & IKZF1	**Panel_2 & IKZF2**	**Panel_2 & IKZF3**	**Panel_2 & IKZF IC**
Panel_3 & IKZF1	**Panel_3 & IKZF2**	**Panel_3 & IKZF3**	**Panel_3 & IKZF IC**
Panel_4 & IKZF1	**Panel_4 & IKZF2**	**Panel_4 & IKZF3**	**Panel_4 & IKZF IC**

***To detect the expression of OX40L on different cell types. Different isotype control cannot be added into the same tube. Therefore, we can add PE isotype control to any of ZF1-3, except for ZF IC.**

22. Wash the cells in 1ml PBS and spin at 350xg for 5 minutes at 4°C.

23. Dilute 1333ul of fixation buffer to 1X concentration using 2667ul **ddH_2O**. **Mix well.**

24. Decant the supernatant from the samples and place them in a tube rack at room temperature. Add 200ul of fixation buffer to each sample and **quickly vortex to fully resuspend the pellet**. Incubate at room temperature in the dark for 10 minutes.

25. Wash the cells in 1ml PBS and spin at 500xg for 5 minutes at 4°C.

C. Intracellular Staining

26. Prepare the 4 master mixes for the intracellular staining. **Add 4ul of the IKZF antibodies, or 2ul of the isotype control, to 400ul of permeabilization buffer in separate tubes**.

27. Decant the supernatant from the samples, ensuring that there is as little supernatant left as possible but without detaching the cell pellet from the tube.

28. **Add 100ul of "blank" permeabilization buffer to the US and the VD SS samples, and 100ul of the IKZF master mixes to their respective samples.** Vortex all tubes to resuspend and incubate on ice in the dark for 20 minutes.

Panel_1 & IKZF1	**Panel_1 & IKZF2**	Panel_1 & IKZF3	**Panel_1 & IKZF IC**
Panel_2 & IKZF1	**Panel_2 & IKZF2**	Panel_2 & IKZF3	**Panel_2 & IKZF IC**
Panel_3 & IKZF1	**Panel_3 & IKZF2**	Panel_3 & IKZF3	**Panel_3 & IKZF IC**
Panel_4 & IKZF1	**Panel_4 & IKZF2**	Panel_4 & IKZF3	**Panel_4 & IKZF IC**

29. Wash the cells in 1ml of 2% PBS/BSA and spin at 500xg for 5 minutes at 4°C. The US and VD SS samples do not need to be washed, leave them in the perm buffer until the final wash.

30. Prepare the master mix for the secondary antibody. Add 10ul of Alexa Fluor 488 anti-goat IgG to 2ml permeabilization buffer. Mix well and place on ice in the dark.

31. Decant the supernatant from the samples and place on ice. Add 100ul of the permeabilization buffer containing the secondary antibody to the other samples

(except the US and VDSS). Vortex all tubes to resuspend and incubate on ice in the dark for 20 minutes.

32. Wash the cells in 1ml of 2% PBS/BSA and spin at 500xg for 5 minutes at 4°C.

33. Decant the supernatant and resuspend in 500ul of 2% PBS/BSA. Place lids on all the tubes and put them in a polystyrene tube holder. Wrap the whole tube holder in foil and keep in the fridge.

Surface Marker Master Mixes (*Mix by pipetting)

Panel 1

178ul	2% PBS/BSA
4ul	CD8 Alexa 700
4ul	**CD4 PE/Texas Red**
4ul	CD45RA BV785
2ul	**CD38 APC**
4ul	CD3 APC/Cy7
4ul	HLA-DR PerCP/Cy5.5

Panel 2

152ul	2% PBS/BSA
20ul	CD25 BV605
4ul	**CD4 PE/Texas Red**
8ul	CD45RO BV785
8ul	CD127 PE/Cy7
4ul	CD3 APC/Cy7
4ul	HLA-DR PerCP/Cy5.5

Panel 3

126ul	2% PBS/BSA
12ul	CD24 PerCP/Cy5.5
4ul	CD19 BV421
8ul	CD27 BV785
2ul	**CD38 APC**
4ul	CD20 PE/Cy7
4ul	CD3 APC/Cy7
20ul	IgD BV605
20ul	CD138 Alexa 700

Panel 4

160ul	2% PBS/BSA
8ul	CD56 BV605
4ul	CD123 APC
4ul	CD11c PE/CF594
4ul	CD16 BV785
4ul	CD3 APC/Cy7
4ul	CD19 BV421
4ul	CD20 PE/Cy7
4ul	HLA-DR PerCP/Cy5.5
4ul	CD14 Alexa 700

Date Written: 16/07/14 **Version:** 5.0
Amended: 09/02/2015_Lingyan Chen

Acquisition and Analysis of Full Immunophenotyping Samples

Reagents Needed:
1x PBS
2% PBS/BSA
Cytometer Setup and Tracking Beads [*BD Biosciences*]
Mouse IgG Compensation Beads [*BD Biosciences*]

Antibodies Needed:
See document "Full Staining Panel.xlsx".

Notes

- This protocol is customized for the Full Immunophenotyping study involving the investigation of IKZF1-3 and OX40L expression on PBMC subsets. Several aspects of this protocol are not directly transferable to other similar protocols (filter setup, CS&T bead lot, application settings).
- All samples should be kept **on ice and in the dark** as much as possible during sample acquisition.

A. Compensation Bead Preparation

1. Label 26 flow cytometry tubes with each of the antibodies being used, with the exception of the IKZF antibodies. Instead, label a tube simply as "Alexa 488".

CCR7 BV421	CXCR3 V605	CCR6 PE/Cy7	CCR4 APC	OX40L PE
CD8 Alexa 700	CD4 PE/Tex Red	CD45RA BV785	CD38 APC	CD3 APC/Cy7
HLA-DR PerCP/Cy5.5	CD25 BV605	CD45RO BV785	CD127 PE/Cy7	CD24 PerCP/Cy5.5
CD19 BV421	CD27 BV785	CD20 PE/Cy7	IgD BV605	CD138 Alexa 700
CD56 BV605	CD123 APC	CD11c PE/CF594	CD16 BV785	CD14 Alexa 700
Anti-mouse IgG Alexa 488				

2. To each tube, add 100ul of 2% PBS/BSA.

3. **Vortex both vials of compensation beads (positive and negative beads).** Add one drop (~60ul) of the positive and negative beads to each tube**. Quickly vortex to mix.**

4. Add 1ul of each of the antibodies to their corresponding tube and **mix gently** (*gently flick the bottom of the tubes to mix). To the Alexa 488 tube, add 1ul of any Alexa 488-labelled mouse IgG antibody.

5. Incubate for 20 minutes at room temperature in the dark.

6. Add 1ml of 2% PBS/BSA to each tube and then centrifuge for 10 minutes at room temperature at 200xg.

7. Remove the supernatant and resuspend the beads in 500ul 2% PBS/BSA. Place the beads in the ice box to be taken to the Flow Core, together with the prepared samples. And one tube with 400ul PBS and the CST beads.

B. Machine Setup

8. Make sure the Fortessa is turned on and the fluidics is set to 'Standby'. Open FACSDiva and login, then wait for the program to connect to the cytometer.
   (*FACSDiva: Username: Lingyan; Password: lingyan)

9. Open the front-right panel on the cytometer to expose the **red laser detector array**. Remove the filter labelled 670/14 from detector C (APC) and replace it with the filter 670/30**.** Ensure that the filter is inserted the right way round, with the bottom of the writing on the label facing the inside of the detector array. Close the panel door.

10. Open the side drawer on the left of the cytometer by pushing the black button in the centre. When it pops out, turn it anticlockwise and pull open the draw to expose the violet, blue and yellow-green laser detector arrays.

11. On the **blue laser detector array**, remove the filter labelled 670/30 from detector A and replace it with the filter 710/50.

12. On the **yellow-green laser detector array**, remove the mirror placeholder from in front of the filter on detector E and insert the longpass mirror labelled 550LP. Close the draw and push the button to lock.

13. On FACSDiva, open the menu 'Cytometer' and select the option 'CST'. Once the CST window has opened, wait for it to connect to the cytometer.

14. Vortex vial of beads. Prepare the CST beads by adding 400ul PBS to a flow cytometry tube and adding 1 drop of CST beads (Lot 43523). Vortex the beads gently before dispensing.

15. In the Setup Control window, click 'Select Configuration'. From the list, select the configuration labelled 'CP-Blue 6Violet 2NUV 3Red 5YG'. Click 'Set Configuration' then click okay. Make sure that the box before "Load tubes manually" has been ticked.

16. Select the correct bead lot from the dropdown menu (43523).

17. Verify that the baseline for this bead lot is detected, then load the bead tube on to the sample acquisition platform. On the machine, press 'Run' and set the acquisition rate to 'Lo'. In the Setup Control window, press 'Run'.

18. Allow the performance check to run. Once the check is complete, make sure that it has passed. If the check has failed, run it again. Otherwise, troubleshoot the problem with the machine.

19. Go to 'File' and 'Exit' to return to FACSDiva.

20. When the ‘CST Mismatch’ dialog appears, click ‘Use CST Settings’. Open a new experiment to run the samples using the CST settings.

C. Sample Acquisition

21. Open a new experiment. Select Cytometer Settings, and in the Inspector window, delete any unused fluorochromes and replace them with any ones that are needed. The following channels should be selected – **FITC, PE, APC, QDot800, QDot605, Pacific Blue, APC-Cy7, PerCP-Cy5.5, PE-Cy7, PE-TexasRed, Alexa700, UV (Violet)**. Tick the ‘H’ and ‘W’ checkboxes under FSC and SSC. Do not adjust any of the voltages.

22. In the new experiment, right click on ‘Cytometer Settings’, go down to ‘Application Settings’ and click ‘Apply’. From the list of available application settings, select ‘CP-VyseImmunophenotyping’.

23. Create the necessary worksheets and samples. In the machine, set the acquisition rate to ‘Hi’.

24. Set the **SSC** voltage to roughly 220 in order to bring the compensation beads into scale. Record **5000 events** for all the compensation bead samples, but do not run the FACSDiva compensation module, since compensation will be done later on FlowJo.

25. Adjust the FSC and SSC voltages for cells (roughly 280 SSC), and record the samples, including the US and VDSS. Set a gate around the PBMCs, and set this as the stopping gate. Record **10,000 events for US and VDSS** and record **300,000 events per sample** (within the stopping gate).

26. Once all samples have been recorded, export the data as FCS files and upload to the web file server (*username: facsdrive; password: music).

27. Begin the cleaning procedure.

28. In the ‘Cytometer’ menu, select ‘View Configurations’. In the CST window, select the original cytometer configuration and apply it to set the machine back to standard settings. Close the window to return to FACSDiva. Once the cleaning is complete, close FACSDiva.

29. Return the filters to their base configurations (* notes behind computer).